SEMI SUPERVISED MACHINE LEARNING

By

ASHISH SHARMA -061222
DEEP CHANDRA GUPTA -061231

MAY-2010

Submitted in partial fulfillment of the Degree of Bachelor of Technology

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING AND INFORMATION TECHNOLOGY

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY-WAKNAGHAT

CERTIFICATE

This is to certify that the work entitled, "Semi Supervised Machine Learning" submitted by Ashish Sharma (061222) and Deep Chandra Gupta (061231) in partial fulfillment for the award of degree of Bachelor of Technology in Computer Science & Engineering of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Mr. Praveen Kumar Tripathi
(Project supervisor)

Department of Computer Science and Engineering,

Jaypee University of Information Technology,

Waknaghat, Solan-173215, India

ACKNOWLEDGMENT

We would like to extend our sincere gratitude to our project mentor Shri Praveen Kumar Tripathi, who provided us inspiring guidance and gave us an insight into the latest technical development going in the vast field of machine learning. We would also like to thank Shri Suman Saha who took a lot a pain and devoted a considerable part of his precious time to give us a better knowledge of the field, without his support this project would not have been possible.

We would like to thank JUIT, Department of Computer Science & Engineering and all other members who provided us this golden opportunity and supported us to take another great step towards the completion of final year project. We owe our heartiest thanks to Brig. (Retd.)S.P.Ghrera (H.O.D CSE/I.T Department) who always inspired confidence in us to take initiative. He has always been motivating and encouraging.

Deep.

DEEP CHANDRA GUPTA

(061231)

Ashish

ASHISH SHARMA

(061222)

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Semi-supervised learning is a learning paradigm concerned with the study of how computers and natural systems such as humans learn in the presence of both labeled and unlabeled data. Traditionally, learning has been studied either in the unsupervised paradigm (e.g., clustering, outlier detection) where all the data are unlabeled or in the supervised paradigm (e.g., classification, regression) where all the data are labeled. The goal of semi-supervised learning is to understand how combining labeled and unlabeled data may change the learning behavior, and design algorithms that take advantage of such a combination. Semi-supervised learning is of great interest in machine learning and data mining because it can use readily available unlabeled data to improve supervised learning tasks when the labeled data are scarce or expensive. Semi-supervised learning also shows potential as a quantitative tool to understand human category learning, where most of the input is self-evidently unlabeled. In this introductory report, we present some popular semi-supervised learning models, including self-training, mixture models, co-training and multi view learning, graph-based methods, and semi-supervised support vector machines. For each model, we discuss its basic mathematical formulation. The success of semi-supervised learning depends critically on some underlying assumptions. We emphasize the assumptions made by each model and give counterexamples when appropriate to demonstrate the limitations of the different models. In addition, we discuss semi-supervised learning for cognitive psychology.

Unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The acquisition of labeled data for a learning problem often requires a skilled human agent to manually classify training examples. The cost associated with the labeling process thus may render a fully labeled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value.

The main goal of the project is to develop a more effective algorithm for semi-supervised learning, develop a tool for the same. It also aims to implement the same in any real life field like image processing, web mining, multimedia processing, medical, bio-informatics datasets, etc, if allowed by the time constraint.

CHAPTER -1

Machine Learning

1.1 Basics

Machine learning studies computer algorithms for learning to do stuff. We might, for instance, be interested in learning to complete a task, or to make accurate predictions, or to behave intelligently. The learning that is being done is always based on some sort of observations or data. Machine learning is about learning to do better in the future based on what was experienced in the past. The emphasis of machine learning is on automatic methods. In other words, the goal into devise learning algorithms that do the learning automatically without human intervention or assistance. Often we have a septic task in mind, such as spam filtering. But rather than program the computer to solve the task directly, in machine learning, we seek methods by which the computer will come up with its own program based on examples that we provide. Machine learning is a core subarea of artificial intelligence. It is very unlikely that we will be able to build any kind of intelligent system capable of any of the facilities that we associate with intelligence, such as language or vision, without using learning to get there. These tasks are otherwise simply too difficult to solve. Further, we would not consider a system to be truly intelligent if it were incapable of learning since learning is at the core of intelligence. Although a subarea of AI, machine learning also intersects broadly with other fields, especially statistics, but also mathematics, physics, theoretical computer science and more.

Some machine learning systems attempt to eliminate the need for human intuition in data analysis, while others adopt a collaborative approach between human and machine. Human intuition cannot, however, be entirely eliminated, since the system's designer must specify how the data is to be represented and what mechanisms will be used to search for a characterization of the data.

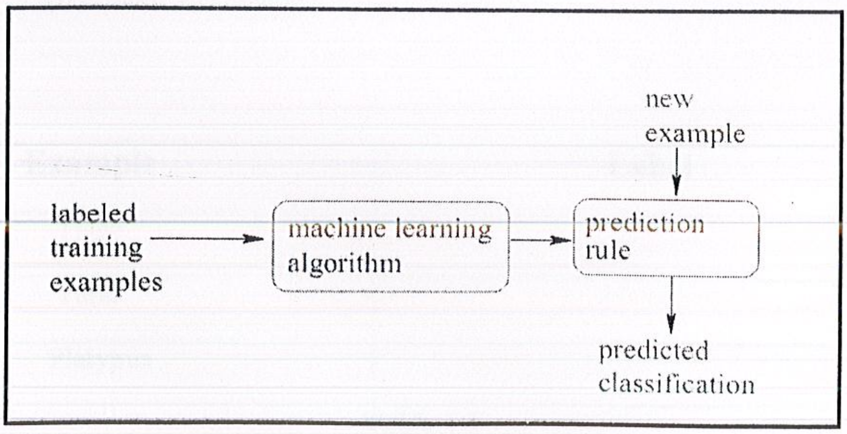


Fig 1.1: A learning problem.

A Tiny Example for Learning:

Training:

Example	Label
Ant	--
Bat	+
Dolphin	--
Leopard	+
Sea lion	--
Zebra	+
Shark	--
Mouse	+
Chicken	--

Table 1.1

Testing:

Example	Label
Tiger	
Tuna	
Platypus	

Table 1.2

Here, examples are labeled positive ("+") or negative ("–"). In this case, the pattern is that "land mammals" are positive, and others negative. The positive examples are animals that don't live in the ocean and don't lay eggs. Thus, test examples "tiger" and "tuna" are positive and negative, respectively, but we can only guess the correct label of "platypus" (an egg-laying mammal).

So in sum, there are three conditions that must be met for learning to succeed.

- First, we need enough data.
- Second, we need to find a rule that makes a low number of mistakes on the training data.
- And third, we need that rule to be as simple as possible.

We sometimes can only find a rule that makes a low number of mistakes by choosing a rule that is more complex, and conversely, choosing a simple rule can sometimes come at the cost of allowing more mistakes on the training data.

1.2 Examples of Machine Learning Problems

There are many examples of machine learning problems. Much of this course will focus on classification problems in which the goal is to categorize objects into a fixed set of categories. Here are several examples:

- Optical character recognition: Categorize images of handwritten characters by the letters represented.
- Face detection: Find faces in images (or indicate if a face is present).
- Spam filtering: Identify email messages as spam or non-spam.
- Tropic spotting: Categorize news articles as to whether they are about politics, sports, entertainment, etc.

- Spoken language understanding: Within the context of a limited domain, determine the meaning of something uttered by a speaker to the extent that it can be classified into one of a fixed set of categories.

- Medical diagnosis: Diagnose a patient as a sufferer or non-sufferer of some disease.

- Customer segmentation: Predict, for instance, which customers will respond to a particular promotion.

- Fraud detection: Identify credit card transactions (for instance) which may be fraudulent in nature.

- Weather prediction: Predict, for instance, whether or not it will rain tomorrow.

1.3 DESIGNING A LEARNING SYSTEM

1.3.1 Choosing the Training Experience

The first design choice we face is to choose the type of training experience from which our system will learn. The type of training experience available can have a significant impact on success or failure of the learner. One key attribute is whether the training experience provides direct or indirect feedback regarding the choices made by the performance system. A second important attribute of the training experience is the degree to which the learner controls the sequence of training examples. A third important attribute of the training experience is how well it represents the distribution of examples over which the final system performance P must be measured. In general, learning is most reliable when the training examples follow a distribution similar to that of future test examples.

1.3.2 Choosing the Target Function

The next design choice is to determine exactly what type of knowledge will be learned and how this will be used by the performance program. Let us begin with a checkers-playing program that can generate the legal moves from any board state. The program needs only to learn how to choose the best move from among these legal moves. This learning task is representative of a large class of tasks for which the legal moves that define some large search space are known a priori, but for which the best search strategy is not known. Many optimization problems fall into this class, such as the problems of scheduling and controlling manufacturing processes where the available manufacturing steps are well understood, but the best strategy for sequencing them is not.

1.3.3 Choosing a Representation for the Target Function

We must choose a representation that the learning program will use to describe the function that it will learn. In general, this choice of representation involves a crucial tradeoff. On one hand, we wish to pick a very expressive representation to allow representing as close an approximation as possible to the ideal target function. On the other hand, the more expressive the representation, the more training data the program will require in order to choose among the alternative hypotheses it can represent.

1.4 PERSPECTIVES AND ISSUES IN MACHINE LEARNING

One perspective on machine learning is that it involves searching a very large space of possible hypotheses to determine one that best fits the observed data and any prior knowledge held by the learner.

1.4.1 Issues in Machine Learning

What algorithms exist for learning general target functions from specific training examples? In what settings will particular algorithms converge to the desired function, given sufficient training data? Which algorithms perform best for which types of problems and representations?

- How much training data is sufficient? What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?
- When and how can prior knowledge held by the learner guide the process of generalizing from examples? Can prior knowledge be helpful even when it is only approximately correct?
- What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem?
- What is the best way to reduce the learning task to one or more function approximation problems? Put another way, what specific functions should the system attempt to learn? Can this process itself be automated?
- How can the learner automatically alter its representation to improve its ability to represent and learn the target function?

CHAPTER-2

Unsupervised Machine Learning

2.1Basics

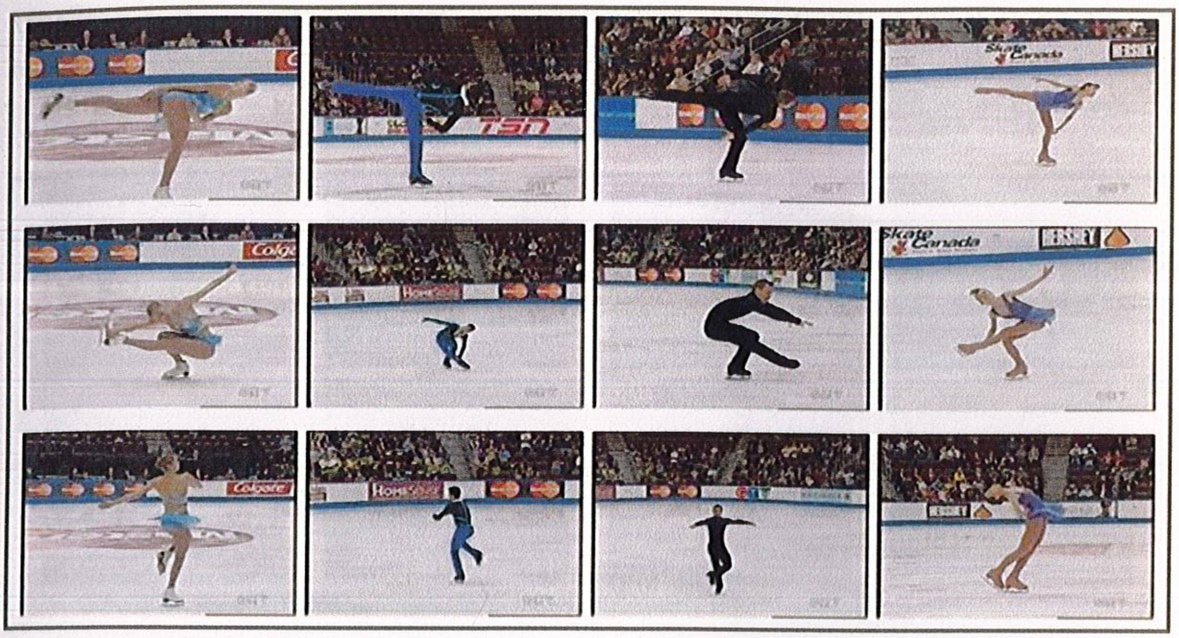
Learning useful structure without labeled classes, optimization criterion, feedback signal, or any other information beyond the raw data is termed as unsupervised learning. There are actually two approaches to unsupervised learning.

The first approach is to teach the agent not by giving explicit categorizations, but by using some sort of reward system to indicate success. Often, a form of reinforcement learning can be used for unsupervised learning, where the agent bases its actions on the previous rewards and punishments without necessarily even learning any information about the exact ways that its actions affect the world. This can be extremely beneficial in cases where calculating every possibility is very time consuming.

On the other hand, it can be very time consuming to learn by, essentially, trial and error. But this kind of learning can be powerful because it assumes no pre-discovered classification of examples.

A second type of unsupervised learning is called clustering. In this type of learning, the goal is not to maximize a utility function, but simply to find similarities in the training data. The assumption is often that the clusters discovered will match reasonably well with an intuitive classification. For instance, clustering individuals based on demographics might result in a clustering of the wealthy in one group and the poor in another.

Although the algorithm won't have names to assign to these clusters, it can produce them and then use those clusters to assign new examples into one or the other of the clusters. This is a data-driven approach that can work well when there is sufficient data; for instance, social information filtering algorithms, such as those that Amazon.com use to recommend books, are based on the principle of finding similar groups of people and then assigning new users to groups.



An example of Semisupervised Data set

Different Acrobats performing similar actions.

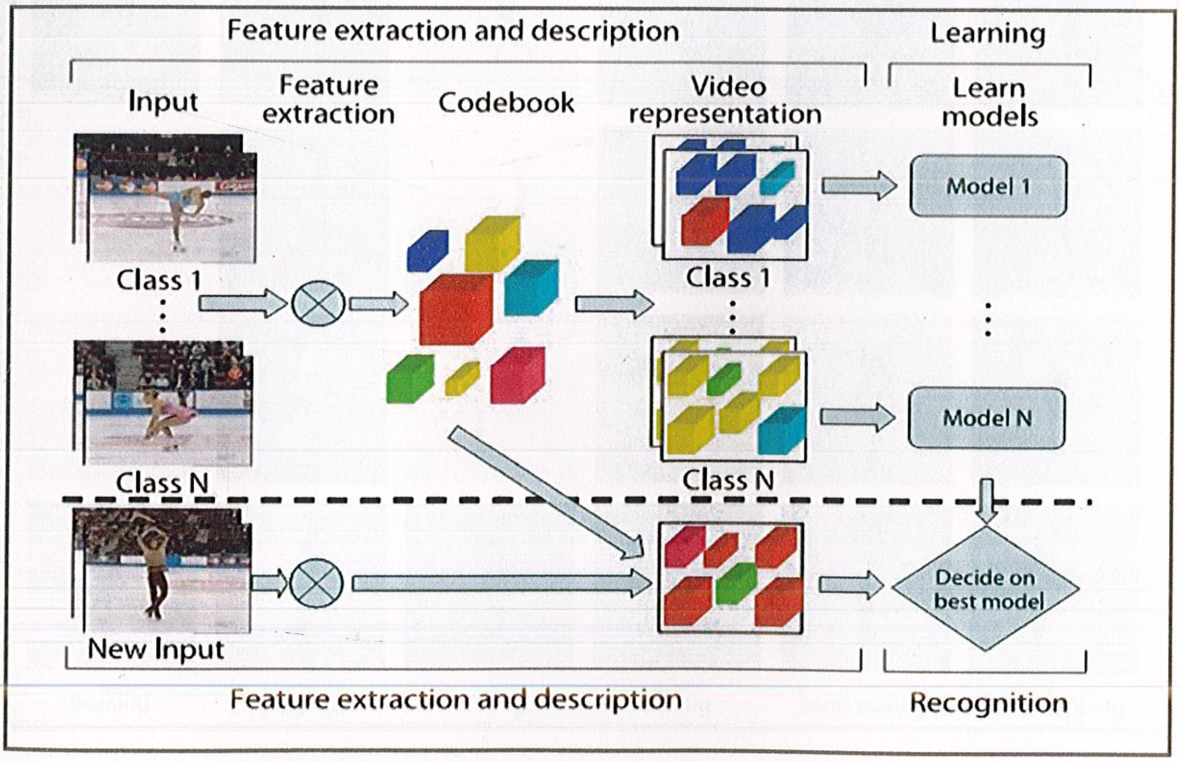
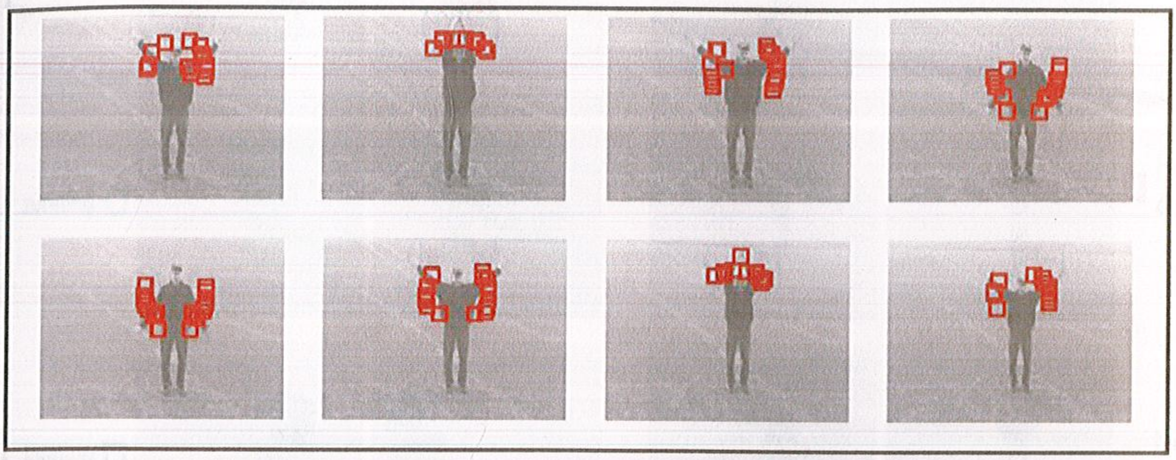


Figure 2.1(a), 2.1(b)

Various red color coordinates in different sets form different usual patterns which can be recognized and can be used for giving different classes.



Various actions tagged according to marked movements.

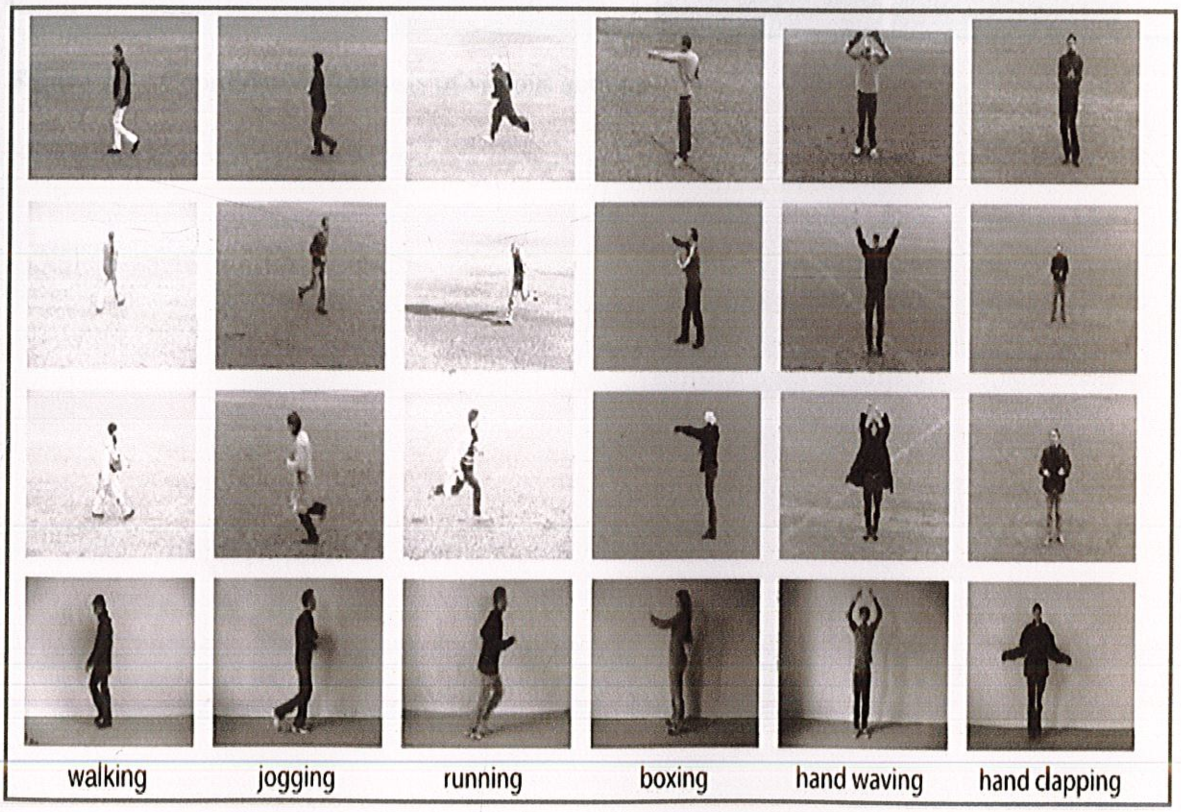


Figure 2.2(a),2.2(b)

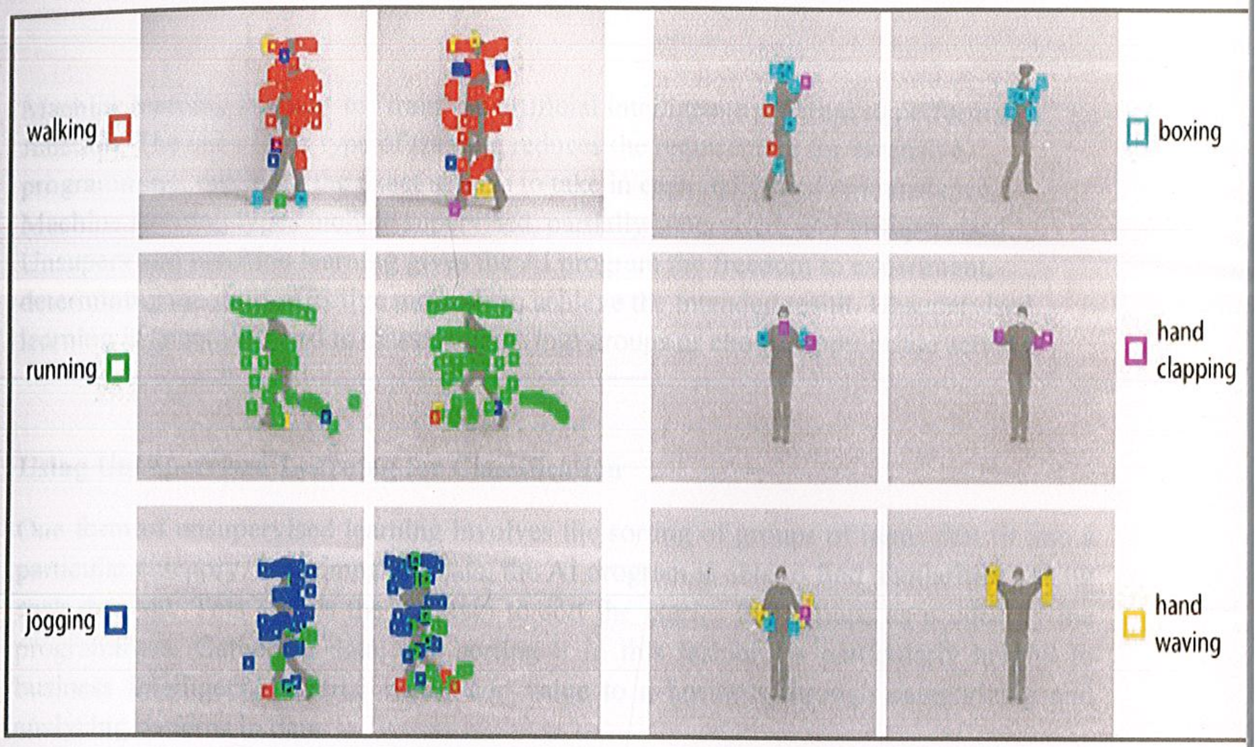


Figure 2.3 : Coordinative Analysis of various actions.

Machine learning is a way to "train" an artificial intelligence program to perform a function. The use of this type of training reduces the requirement for extensive programming, outlining the exact actions to take in each individual circumstance. Machine learning types include supervised, partially supervised, and unsupervised. Unsupervised machine learning gives the AI program the freedom to experiment, determining the most effective methods to achieve the intended result. Unsupervised learning is generally used to classify items into groups or choose appropriate actions.

Using Unsupervised Learning for Classification

One form of unsupervised learning involves the sorting of groups of items that fit into a particular category. By comparing data, the AI program is able to find similarities among each data set. This allows the program to sort the entries into groups as needed by the programmers. Gathering data, and sorting it in this fashion, is particularly helpful to business intelligence efforts, which add value to a business through categorizing and analyzing patterns in data.

Teaching an AI Program to Make Decisions

The other form of unsupervised machine learning allows programmers to teach AI programs how to make good decisions. The AI program receives rewards for correct classification of items during the training phase. The software stores the results of all attempts, including all feedback received. Unsupervised learning methods force the program to base future attempts on past actions, learning from established failures or successes.

2.2 Major Approaches based on Supervised Learning:

- Clustering (n-link, k-means, GAC,...)
- Taxonomy creation (hierarchical clustering)
- Novelty detection ("meaningful" outliers)
- Trend detection (extrapolation from multivariate partial derivatives)

Flow chart for K-Mean algorithm

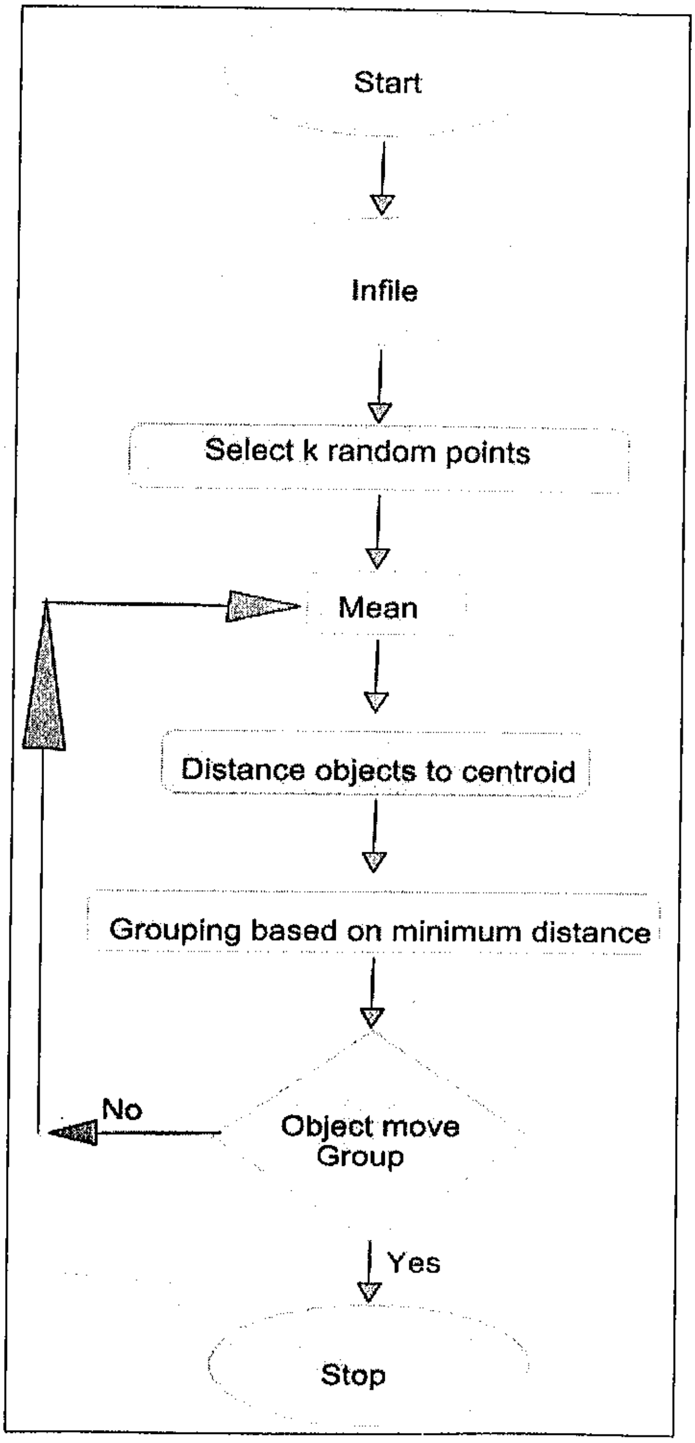


Figure 2.4

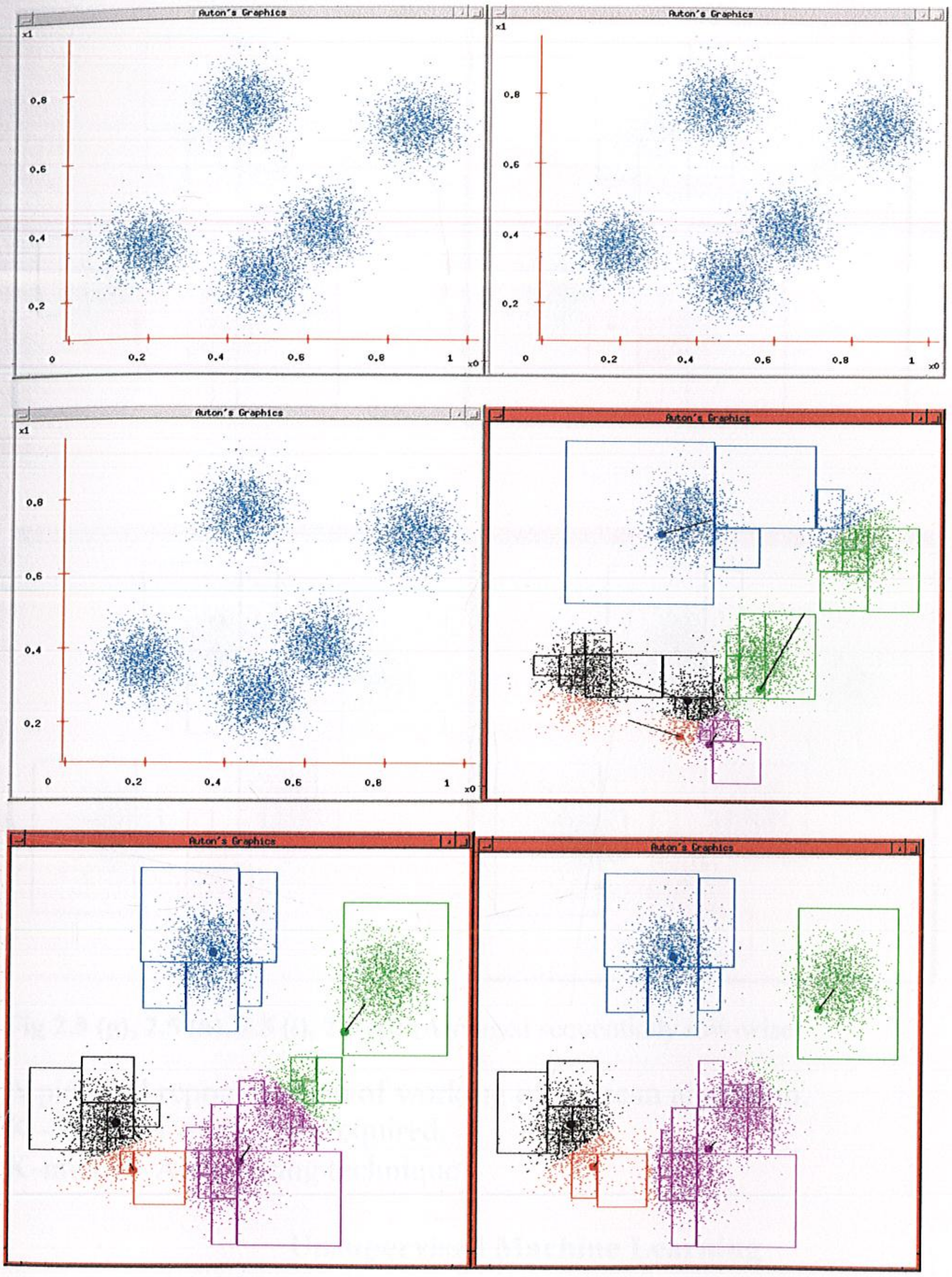


Fig 2.5 (a), 2.5 (b), 2.5 (c), 2.5 (d), 2.5 (e), 2.5 (f) :(Arranged sequentially row-wise)

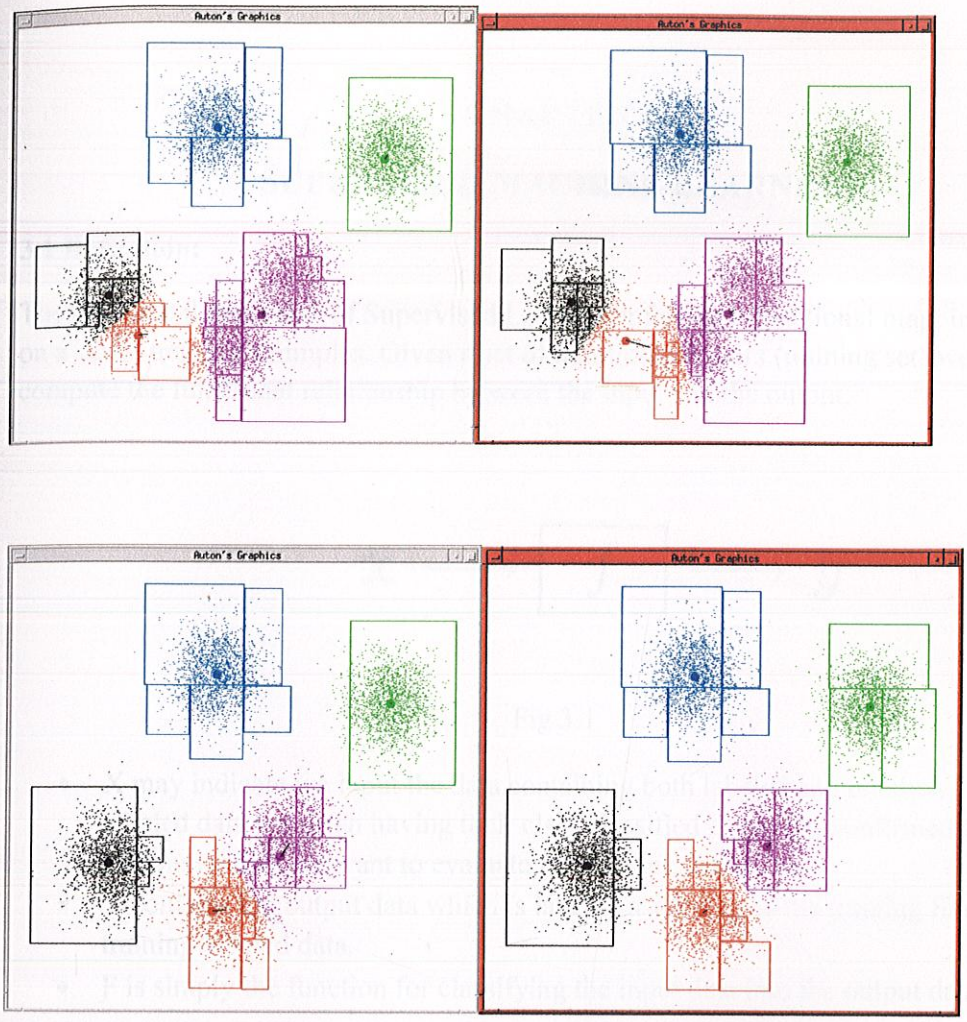


Fig 2.5 (g), 2.5 (h), 2.5 (i), 2.5 (j): (Arranged sequentially row-wise)

A pictorial representation of working of K-mean algorithm.
K - Number of clusters required.
K-mean – A clustering technique

Unsupervised Machine Learning

CHAPTER 3

SUPERVISED MACHINE LEARNING

3.1 Definition:

The theoretical definition of Supervised Learning inferring a functional mapping based on a set of training examples. Given a set of input/output pairs (training set) we wish to compute the functional relationship between the input and the output.

$$\mathbf{X} \longrightarrow \boxed{f} \longrightarrow y$$

Fig 3.1

- X may indicate we input the data containing both labeled and unlabeled data. By labeled data we mean having their class classified or having confirmed their attribute which we want to evaluate for unlabeled data.
- Y contains the output data which is labeled according to the training function of training labeled data.
- Ƒ is simply the function for classifying the input data into the output data.

In simple terms Supervised learning utilizes the labeled data from input data for training data. Labeling here refers that class or final attribute is known. If labeled data is not separately available then some training data is utilized for training function and unlabeled data is classified using the same function of training data of the same sample set.

For e.g. in people detection given an image we wish to say if it depicts a person or not. The output is one of 2 possible categories
In pose estimation we wish to predict the pose of a face image. The output is a continuous number (here a real number describing the face rotation angle).

Input data is generally highly dimensional that means exponential increase in volume associated with adding extra dimensions to a space.

3.2 Learning approach:

Learning attempts to infer the algorithm for a set of (labeled) examples in much the same way that children learn by being shown a set of examples (e.g. sports/non sports car).

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. In other words, the goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features.

Statistical Classification

A typical supervised learning problem is comprised of two components: (1) an outcome measurement, which can be either quantitative or categorical (such as fraudulent/not fraudulent elections); and (2) a training set of data, which includes the outcome and feature measurements for a set of objects (such as electoral contests). The process of applying supervised machine learning to a real-world problem involves a series of steps :

(1) Identification of required data. For any given problem we need a set of variables, denoted as inputs, features, or predictors (i.e. the independent variables), which are measured or preset. These should have some influence on one or more outputs (i.e. the responses or the dependent variables). Thus, a set of input and output objects should be gathered, either from real or simulated data.

(2) Definition of a training set. The training data consist of pairs of input objects (typically vectors), and desired outputs. The outputs may vary in nature; they can take a continuous value (in this case, the prediction task is usually called regression), or they may predict a class label of the input object (and thus, the prediction task received the name of classification).

(3) Determination of the input feature representation. The accuracy of the supervised learner depends strongly on how the input object is represented. The input vector should contain a number of features that are descriptive of the input object. In addition, the feature selection process should identify and remove as many irrelevant and redundant features as possible. This reduces the dimensionality of the data and enables the learning process to operate faster.

(4) Algorithm selection. The choice of a specific learning algorithm is a critical step. The task of the learning algorithm is to produce a classifier (hypothesis, function) to classify unlabeled objects into the correct class. A wide range of learning methods are available. These include rule-based learning systems (decision trees, one rule, decision rules), statistical learning systems (naive bayes, support vector machines, artificial neural networks), and ensemble methods (stacking, bagging and boosting). Each of these learning algorithms has its strengths and weaknesses. The main factor in choosing a learning method is its prediction accuracy. However, there is no single classifier that works best on all given problems, and classifier performance depends greatly on the characteristics of the data to be classified.

(5) Model Assessment. The last step is to evaluate the performance of the classifier on new data. The usual approach is to divide the dataset into two parts: a training set, and a test set. The training set should be used to evaluate the performance of different models (input representation and learning algorithm), and to choose most appropriate one; the test set should be used for assessment of the generalization error of the final chosen model. Ideally, the test set should be brought out only at the end of the data analysis

In many machine learning problem domains large amounts of data are available but the cost of correctly labeling it prohibits its use. This paper presents a short overview of methods for using a small set of labeled data together with a large supplementary unlabeled dataset in order to learn a better hypothesis than just by using the labeled information.

In the recent years, enormous amounts of information has become available most notably unstructured and semi -structured textual data available from the internet. In order for this information to be of greater use, more structure needs to be discovered in it to enable automated processing and reasoning. One of the tools used for this is machine learning. Supervised machine learning is a process of learning a function based on given examples. The examples are provided as ordered pairs of objects (A, B) and the learning algorithm induces the function f: A ‡ B based on some inductive bias (prior knowledge / assumptions) which is needed for any generalization to be possible. The resulting function can then be used to map objects into unknown target values.

Since the data available can have a large complexity this inherently means complex functions to be learned and learning complex functions requires many examples. Examples with known target values (a.k.a. labeled examples) are however usually not directly available and need to be manually created, which can be a time-consuming and/or expensive process. In order to minimize this cost, a lot of research has been conducted in the area of using unlabeled examples to aid in the process. Two different approaches and their mixtures will be presented here; one designed to minimize required human effort and the other to work with a fixed set of labeled and unlabeled examples.

The performance of the aforementioned techniques (measured by the expert labeling cost) can vary from problem to problem by orders of magnitude. Computational cost should also be taken into account when choosing an approach while decreasing the cost of human labor the CPU requirements can increase beyond any reasonable limit: in the usual learning scenario one only needs to train one model which can already be an expensive procedure. For a simple uncertainty-based active learning, one has to train the same number of models as there are labeled examples at the end and use every one of them to test each unlabeled sample.

It is possible to decrease the amount of CPU work by a constant factor at the expense of some human labor by selecting several examples at the same time without updating the rest of the system. For the method based on SVM margin sizes, the number of trained and discarded models is for each iteration of active learning loop linearly dependant on the size of the unlabeled set; making efficient implementation of incremental learning algorithms an absolute must.

Pre-processing

Tree-based methods do not require much pre-processing, except that the data must be storable in a standard attribute-object table. Numerical input variables do not need to be normalized and categorical input variables can be treated as such. Although it is always a good idea to remove irrelevant variables, tree-based methods are quite robust to their presence and there is thus usually no need to filter variables prior to building the model. With the goal of obtaining the last percent of accuracy, it is however often possible to get some improvement by applying some feature selection techniques to remove these irrelevant variables (using for example tree-based attribute importance measures themselves) prior to building a predictive model with the tree-based supervised learning methods.

Like other machine learning methods, decision trees typically suffer in the presence of unbalanced datasets, i.e., datasets which contain many more examples from one class than of the others. Although sophisticated approaches exist to deal with this problem, a simple and often appropriate way to handle it is to merely under-sample the majority class so that it contains a comparable amount of examples with respect to the other classes.

Supervised Learning Algorithms

One-rule is the easiest way to find very simple classification rules from a set of instance, it generates a one-level decision tree expressed in the form of a set of rules that all test one particular attribute. Naïve Bayes method is based on Bayesian theorem and a naïve assumption that all attributes are equally important and independent of one another for the given class. It estimates the Gaussian distribution of the attributes for each class based on pre-labeled training set, and then uses the prior probability to determine a new instance's posterior probability. The kernel estimation addresses the problem of approximating every attribute by a Gaussian distribution.

Multilayer perception is a classification scheme based on neural network. The architecture consists of a single input layer of the features, a single output layer of the classes and one or more hidden intervening layers that comprising a number of nodes. These nodes are connected to all nodes in adjacent layers with different weights. The algorithm tunes the weights according to all the training instances, the problem is analogous to that of fitting a function to a set of data

Decision tree is a model based on a tree structure with nodes representing features and branches representing possible values connecting the features. It uses divide-and-conquer approach to build a tree with entropy based gain ratio splitting criterion. Random forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The algorithm has been shown to have desirable properties such as convergence of generalization errors.

AdaBoost is an advanced boosting algorithm which attempts to increase the accuracy of any given learning algorithm.

Feature Selection

Feature selection and dimension reduction plays a very important role in machine learning. We apply feature selection algorithms to the flow feature set for two-fold benefits: the reduction in the number of flow features decreases learning and classification times, while the removal of irrelevant or redundant features can also increase the classification accuracy and help to identify the most important features of the network traffic. Feature selection algorithms are broadly categorized into filter and wrapper . Filter methods analyze the characteristics of the training data to determine the relevance and importance of certain features to the classification problem. On the other hand, wrapper methods evaluate the performance of a classifier using different subsets of features hence obtain the optimal subset that are suited for a particular classifier. Since the purpose of this study is to compare the performance of various classifiers, we use filter methods. There are two classes of filter algorithms: ranking algorithms that provide a goodness measure for individual features and subset search algorithms that pro-vide a goodness measure for subsets of features. We apply two algorithms of each class: consistency-based subset search (CON), correlation-based feature selection (CFS), Chi-squared ranking (CHI) and gain ratio ranking (GR). And the candidate subsets are generated from the feature space using best-first search.

Evaluation Methods

To test the statistical machine learned network traffic classifiers, we use totally different datasets for training and testing. To measure the effectiveness of the classifiers, we adopt three metrics that are widely used in the data mining literature accuracy, precision and recall. Accuracy is the percentage of correctly classified flow instances over the total number of instances. Precision is the number of class members classified correctly over the total number of instances classified as class members. Recall is the number of class members classified correctly over the total number of class members.

People detection example

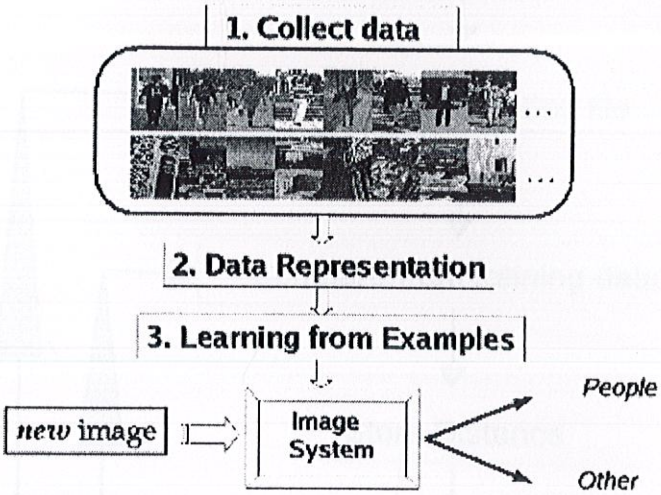


Fig 3.2

3.3 Major Approaches based on Supervised Learning:

- Decision Tree
- Naive Bayesian classifier
- Nearest Neighbor

•Naive Bayesian classifier: The Naive Bayesian Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayesian can often outperform more sophisticated classification methods.

3.4 Applications:

• Optical digit recognition (useful for identifying the numbers in a ZIP code from a digitalized image) (Computer Vision)

• Predicting house prices based on sq. feet, number of rooms, distance from central London,... (Marketing)

• Estimate amount of glucose in the blood of a diabetic person (Medicine)

• Detect spam emails (Information retrieval)

• Predict protein functions / structures (Bioinformatics)

• Speaker identification / sound recognition (Speech recognition)

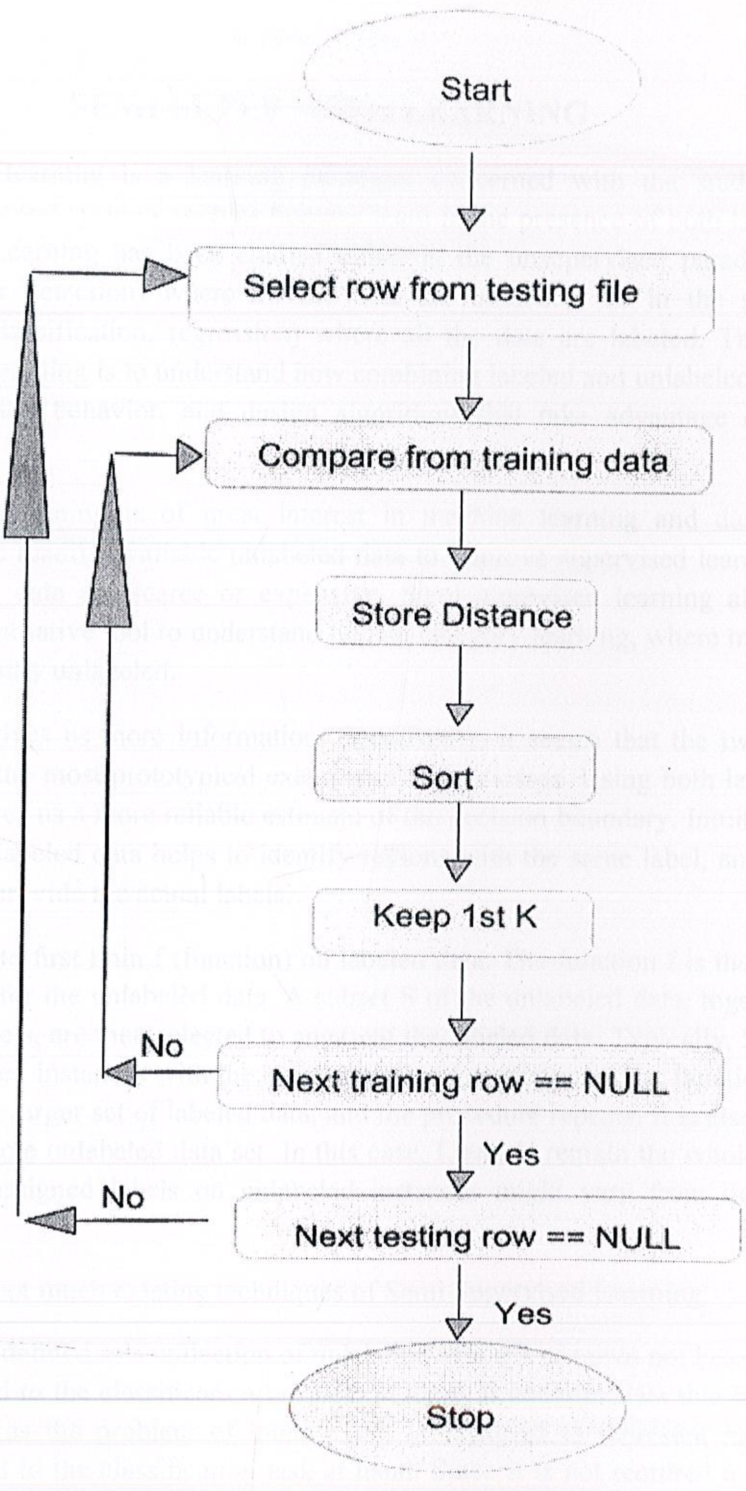

Start
Select row from testing file
Compare from training data
Store Distance
Sort
Keep 1st K
No
Next training row == NULL
Yes
No
Next testing row == NULL
Yes
Stop

Fig 3.3: Flowchart of KNN algorithm

CHAPTER 4

SEMI-SUPERVISED LEARNING

Semi-supervised learning is a learning paradigm concerned with the study of how computers and natural systems such as humans learn in the presence of both labeled and unlabeled data. Learning has been studied either in the unsupervised paradigm (e.g., clustering, outlier detection) where all the data are unlabeled or in the supervised paradigm (e.g., classification, regression) where all the data are labeled. The goal of semi-supervised learning is to understand how combining labeled and unlabeled data may change the learning behavior, and design algorithms that take advantage of such a combination.

Semi-supervised learning is of great interest in machine learning and data mining because it can use readily available unlabeled data to improve supervised learning tasks when the labeled data are scarce or expensive. Semi-supervised learning also shows potential as a quantitative tool to understand human category learning, where most of the input is self-evidently unlabeled.

Unlabeled data gives us more information. Specifically, it seems that the two labeled instances are not the most prototypical examples for the classes. Using both labeled and unlabeled data gives us a more reliable estimate of the decision boundary. Intuitively, the distribution of unlabeled data helps to identify regions with the same label, and the few labeled data then provide the actual labels.

The main idea is to first train f (function) on labeled data. The function f is then used to predict the labels for the unlabeled data. A subset S of the unlabeled data, together with their predicted labels, are then selected to augment the labeled data. Typically, S consists of the few unlabeled instances with the most confident f predictions. The function f is re-trained on the now larger set of labeled data, and the procedure repeats. It is also possible for S to be the whole unlabeled data set. In this case, L and U remain the whole training sample, but the assigned labels on unlabeled instances might vary from iteration to iteration.

As such there are not much existing techniques of Semi Supervised Learning.

The Universum is defined as a collection of unlabeled examples known not belong to any class that is related to the classification problem at hand. It contains data that belongs to the same domain as the problem of interest and is expected to represent meaningful information related to the classification task at hand. Since it is not required to have the same distribution with the training data, the Universum can reveal some prior information for the possible classifiers. Thishas been justified on inductive classification problems by the Universum support vector machine (U -SVM)

4.1 Semi-Supervised Learning Paradigms

Naively, given the definition above, one could see semi-supervised learning as a way of mixing unsupervised learning and supervised learning together. So for example one could perform density estimation or clustering on the unlabeled data and classification on the labeled data, somehow combining these predictions into a shared model. Indeed, such combinations are possible, and several variants already exist. In the following paragraphs we describe three particular paradigms for semi-supervised learning that fall under this description.

Supervised setting: The typical setting for measuring the success of semi-supervised learning is to treat it as a supervised learning problem where one has additional unlabeled data that could potentially be of benefit. One first trains a model to predict labels given the labeled training set and unlabeled data, and then measures the error rate on a separate labeled test set. However, other interpretations of semi-supervised learning are possible.

Unsupervised setting: One can also be interested in the task of unsupervised learning where one has additional labeled data that could potentially be of benefit. So for example, one can learn a clustering of the data, where one is given some extra must-link or must not-link constraints that are implicit in the labeled data. This paradigm has the advantage of being able to handle missing classes, which could be important in some problems, e.g. speaker identification or protein family detection, to name two.

Level-of-detail setting: Finally, yet another way of looking at semi-supervised learning is to see training labels yi for examples xi as having various levels of detail (granularity). For example in text processing, at a very coarse level one could label whether example sentences are grammatically correct or not. However, a labeling with a finer level of detail would also label the parts-of-speech of the words in the sentence. An even finer level of detail could specify the syntactic parse tree of the sentence. Seen this way, semi-supervised learning should handle labels yi that go between two extremes: from no label at all to very detailed label information.

Each example can be labeled with a different detail-level and the learning algorithm has to be designed to deal with this fact.

Why Semi-Supervised Learning?

Supervised data is expensive both in monetary cost and labeling time. Labeling sentences with parse trees or finding protein function or 3D structure, to give two examples of labeled data, require human expertise. Unlabeled data, on the other hand, is cheap to collect in many domains: audio-based problems, vision problems, and text processing problems, to name a few. In fact, most sensory-based problems those humans are good at have an abundant supply of unlabeled data. However, there are also other kinds of data, not natural to a human's perception, which are also relatively easy to collect compared to labeled examples of that same data. So, returning to our bioinformatics example, knowing the function (label) of a protein is costly, but obtaining its unlabeled primary structure (sequence of amino acids) is cheap. In our view, true AI that mimics humans would be able to learn from a relatively weak training signal. For example, in the field of natural language processing, linguists label sentences with parse trees, but humans learn from data which usually has significantly less detailed labels. This argument perhaps strengthens the importance of semi-supervised learning as a topic of study in the field of machine learning.

4.2 How Does Unlabeled Data Help in a Supervised Task?

Unlabeled data somehow gives knowledge about the density p(x) but tells you nothing about the conditional density one is interested in p(y|x) unless some assumptions are made in a training algorithm that hold true for a particular dataset. There are several possible assumptions that one can make about the data, each leading to different algorithms.

The cluster assumption One can assume that examples in the same cluster have the same class label. This implies that one should perform low density separation; that is, the decision rule one constructs should lie in a region of low density.

The manifold assumption One can also assume that examples in the same manifold have the same class. This is somewhat similar to the cluster assumption, but motivates different algorithms.

Zipf 's law effect One obvious way unlabeled data can help in language problems is that one gets to see words that one has never seen before because a finite training set cannot cover the language.

Zipf's law states that in a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table. In language problems, we are effectively always seeing new features (i.e. new words) every time we look at new examples.

Our conjecture is that effective use of this knowledge in algorithms should surely improve performance over supervised learning alone.

Non-i.i.d. data Many machine algorithms and toy datasets assume that data are identically and independently distributed (i.i.d.) but in real life data this may not be true. We conjecture that if you see a test set that is drawn from a (slightly) different distribution to the training set then semi-supervised learning might help compared to purely supervised learning which is unaware of the distribution of the test set. For example, one might train a parser on the Wall Street Journal, but then apply it to the novel Moby Dick.

4.3 Large-Scale Semi-Supervised Learning Algorithms

Graph-Based Approaches

If one performs early stopping of the iterative algorithm for label propagation, instead of a direct optimization of the given objective function, one can speed this algorithm up. Such an approximation might give a loss of accuracy in the final predictions.

Change of Representation Approaches

For change of representation methods, the problem has been split into two tasks: an unsupervised learning algorithm and a supervised learning algorithm. Clearly then in order to perform large-scale semi-supervised learning, one should choose algorithms from both of these two areas that also scale well.

Margin-Based Regularization Approaches

Semi-supervised learning is useful when labels are expensive, when unlabeled data is cheap and when p(x) is useful for estimating p(y|x), e.g. if either the manifold or cluster assumptions are true. We have reviewed several different algorithmic techniques for encoding such assumptions into learning; generally this is done by somehow "marrying" unsupervised learning into a supervised learning algorithm. Instances of this approach are graph-based approaches, change of representation based approaches and margin based approaches. All of these can somehow be seen as either explicitly or implicitly adding a regularizer that encourages that the chosen function reveals structure in the unlabeled data.

Large-scale learning is often realistic only in a semi-supervised setting because of the expense of labeling data. Moreover, the utility of an unlabeled example is less than a labeled one, thus requiring a relatively large collection of unlabeled data for its use to be effective. However, to make current algorithms truly large-scale, probably only linear complexity with respect to the number of examples will suffice. At least in the non-linear case, current approaches still fall short, leaving the field still open for further research.

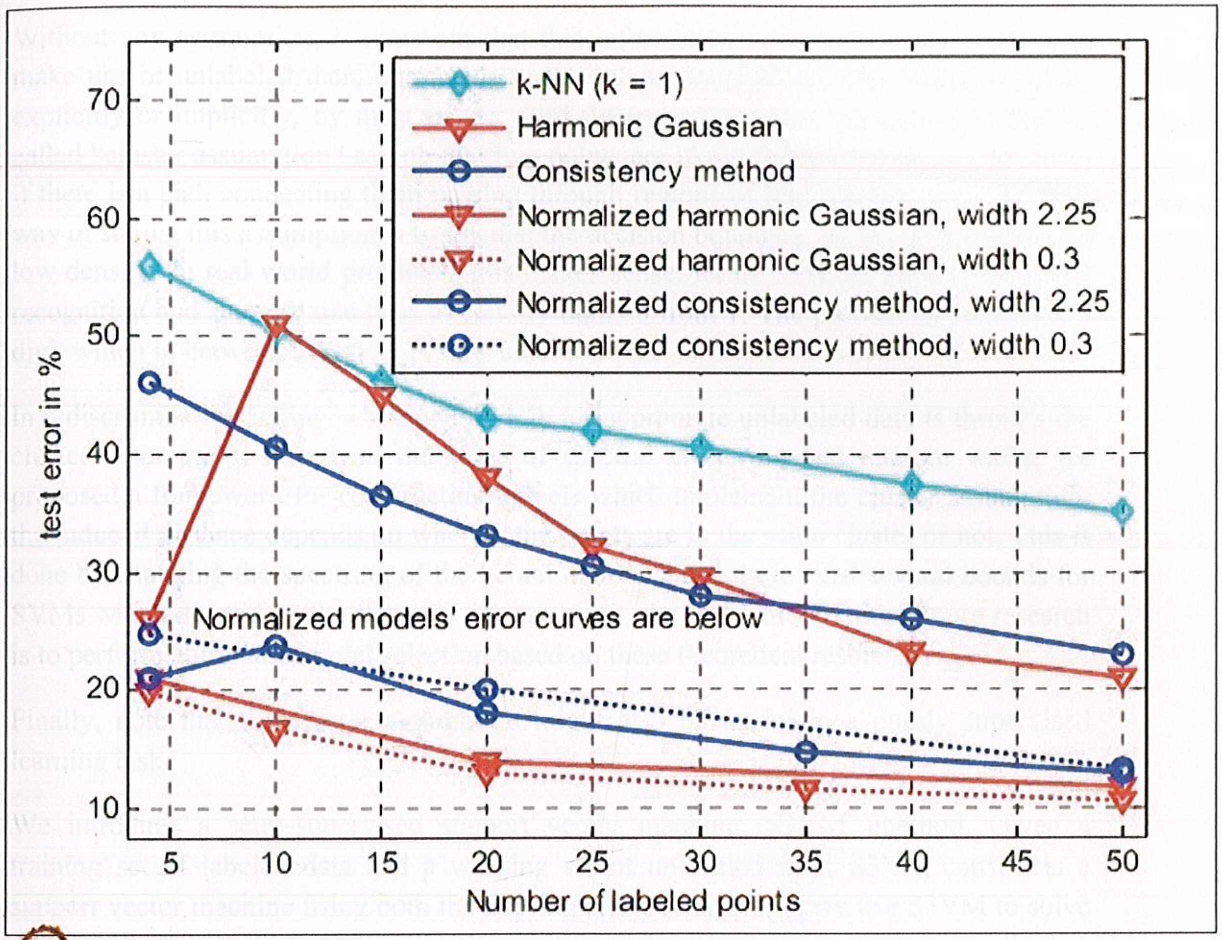


Figure 4.1

We consider the problem of semi-supervised learning, where one has usually few labeled examples and a lot of unlabeled examples. One of the first semi-supervised algorithms was applied to web page classification. This is a typical example where the number of unlabeled examples can be made as large as possible since there are billions of web page, but labeling is expensive since it requires human intervention. Since then, there has been a lot of interest for this paradigm in the machine learning community; an extensive review of existing techniques can be found .

It has been shown experimentally that under certain conditions, the decision function can be estimated more accurately, yielding lower generalization error. However, in a discriminative framework, it is not obvious to determine how unlabeled data or even the perfect knowledge of the input distribution P(x) can help in the estimation of the decision function.

Without any assumption, it turns out that this information is actually useless. Thus, to make use of unlabeled data, one needs to formulate assumptions. One which is made, explicitly or implicitly, by most of the semi-supervised learning algorithms is the so-called "cluster assumption" saying that two points are likely to have the same class label if there is a path connecting them passing through regions of high density only. Another way of stating this assumption is to say that the decision boundary should lie in regions of low density. In real world problems, this makes sense: let us consider handwritten digit recognition and suppose one tries to classify digits 0 from 1. The probability of having a digit which in between a 0 and 1 is very low.

In a discriminative setting, a reasonable way to incorporate unlabeled data is through the cluster assumption. Based on the ideas of spectral clustering and random walks, we proposed a framework for constructing kernels which implement the cluster assumption: the induced distance depends on whether the points are in the same cluster or not. This is done by changing the spectrum of the kernel matrix. Since there exist several bounds for SVMs which depend on the shape of this spectrum, the main direction for future research is to perform automatic model selection based on these theoretical results.

Finally, note that the cluster assumption might also be useful in a purely supervised learning task.

We introduce a semi-supervised support vector machine (S3VM)method. Given a training set of labeled data and a working set of unlabeled data, S3VM constructs a support vector machine using both the training and working sets. We use S3VM to solve the transduction problem using overall risk minimization (ORM) posed by Vapnik. The transduction problem is to estimate the value of a classification function at the given points in the working set. This contrasts with the standard inductive learning problem of estimating the classification function at all possible values and then using the fixed function to deduce the classes of the working set data. We propose a general S3VM model that minimizes both the misclassification error and the function capacity based on all the available data.

We show how the S3VM model for 1-norm linear support vector machines can be converted to a mixed-integer program and then solved exactly using integer programming. Results of S3VM and the standard 1-norm support vector machine approach are compared on eleven data sets. Our computational results support the statistical learning theory results showing that incorporating working data improves generalization when insufficient training information is available. In every case, S3VM either improved or showed no significant difference in generalization compared to the traditional approach.

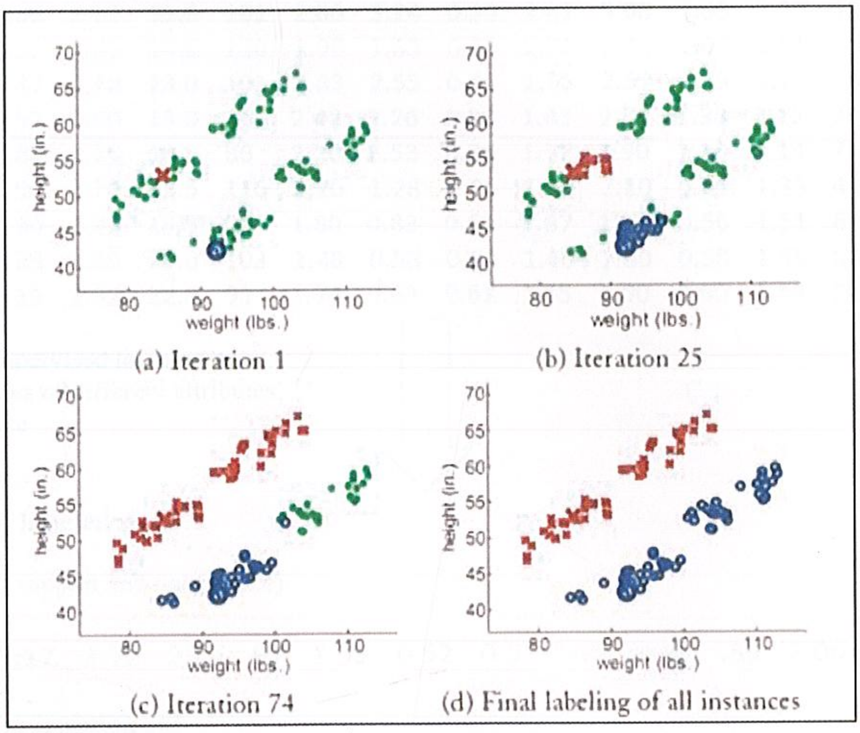
In classification, the transduction problem is to estimate the class of each given point in the unlabeled working set. The usual support vector machine (SVM) approach estimates the entire classification function using the principle of statistical risk minimization (SRM). In transduction, one estimates the classification function at points within the working set using information from both the training and working set data. Theoretically, if there is adequate training data to estimate the function satisfactorily, then SRM will be sufficient. We would expect transduction to yield no significant improvement over SRM alone. If, however, there is inadequate training data, then ORM may improve generalization on the working set. Intuitively, we would expect ORM to yield improvements when the training sets are small or when there is a significant deviation between the training and working set subsamples of the total population.

We introduced a semi-supervised SVM model. S3VM constructs a support vector machine using all the available data from both the training and working sets. We show how the S3VM model for 1-norm linear support vector machines can be converted to a mixed-integer program. One great advantage of solving S3VM using integer programming is that the globally optimal solution can be found using packages such as CPLEX. Using the integer S3VM we performed an empirical investigation of transduction using overall risk minimization, a problem posed by Vapnik. Our results support the statistical learning theory results that incorporating working data improves generalization when insufficient training information is available. In every case, S3VM either improved or showed no significant difference in generalization compared to the usual structural risk minimization approach.

Our empirical results combined with the theoretical results in , indicate that transduction via ORM constitutes a very promising research direction.

Some SSL Techniques:

1. Self Training.
2. Generative Models.
3. S3VMs.
4. Graph-Based Algorithms Multiview Algorithms.



(a) Iteration 1

(b) Iteration 25

(c) Iteration 74

(d) Final labeling of all instances

Fig 4.2

Self Training

Basic Steps in Self Training:

1. Train *f* from (X_l, Y_l); 2. Predict on $x \in X_u$
2. Add (x, *f*(x)) to labeled data
3. Repeat

It is the simplest SSL method and can be easily implemented on the complex existing classifiers. It can be easily implemented on the current trend's real time data.

Training data set

1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
2	12.33	0.99	1.95	14.8	136	1.90	1.85	0.35	2.76	3.40	1.06	2.31	750
2	12.7 0	3.87	2.40	23.0	101	2.83	2.55	0.43	1.95	2.57	1.19	3.13	463
2	12.00	0.92	2.00	19.0	86	2.42	2.26	0.30	1.43	2.50	1.38	3.12	278
2	12.72	1.81	2.20	18.8	86	2.20	2.53	0.26	1.77	3.90	1.16	3.14	714
3	13.11	1.90	2.75	25.5	116	2.20	1.28	0.26	1.56	7.10	0.61	1.33	425
3	13.23	3.30	2.28	18.5	98	1.80	0.83	0.61	1.87	10.52	0.56	1.51	675
3	12.58	1.29	2.10	20.0	103	1.48	0.58	0.53	1.40	7.60	0.58	1.55	640
3	13.17	5.19	2.32	22.0	93	1.74	0.63	0.61	1.55	7.90	0.60	1.48	725

Semi-supervised learning of
properties of different attributes
classwise

Generation of labeled data

(On the basis of support and confidence)

3 13.52 3.17 2.72 23.5 97 1.55 0.52 0.5 0.55 4.35 .89 2.06 520

Adding data to training set

1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
2	12.33	0.99	1.95	14.8	136	1.90	1.85	0.35	2.76	3.40	1.06	2.31	750
2	12.7 0	3.87	2.40	23.0	101	2.83	2.55	0.43	1.95	2.57	1.19	3.13	463
2	12.00	0.92	2.00	19.0	86	2.42	2.26	0.30	1.43	2.50	1.38	3.12	278
2	12.72	1.81	2.20	18.8	86	2.20	2.53	0.26	1.77	3.90	1.16	3.14	714
3	13.11	1.90	2.75	25.5	116	2.20	1.28	0.26	1.56	7.10	0.61	1.33	425
3	13.23	3.30	2.28	18.5	98	1.80	0.83	0.61	1.87	10.52	0.56	1.51	675
3	12.58	1.29	2.10	20.0	103	1.48	0.58	0.53	1.40	7.60	0.58	1.55	640
3	13.17	5.19	2.32	22.0	93	1.74	0.63	0.61	1.55	7.90	0.60	1.48	725
3	13.52	3.17	2.72	23.5	97	1.55	0.52	0.50	0.55	4.35	0.89	2.06	520

Data added to training set along with label

CHAPTER 5

Fig 5.1
Level 0 DFD

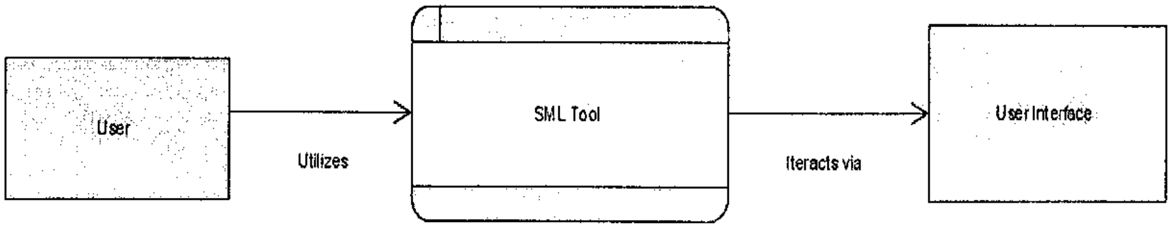
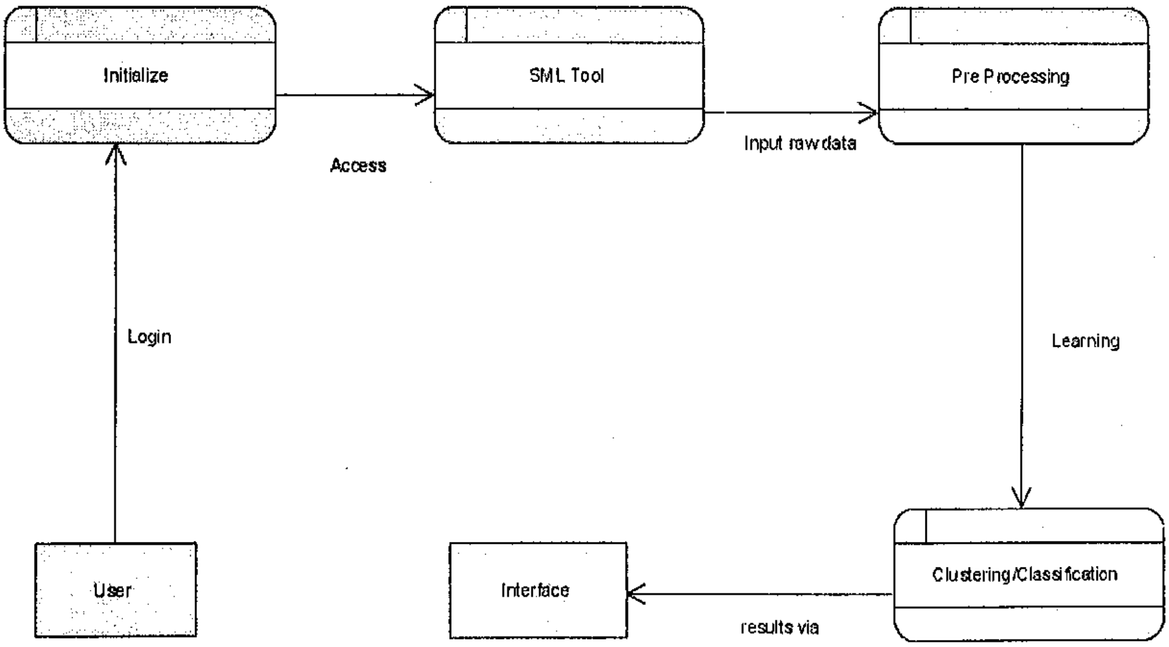
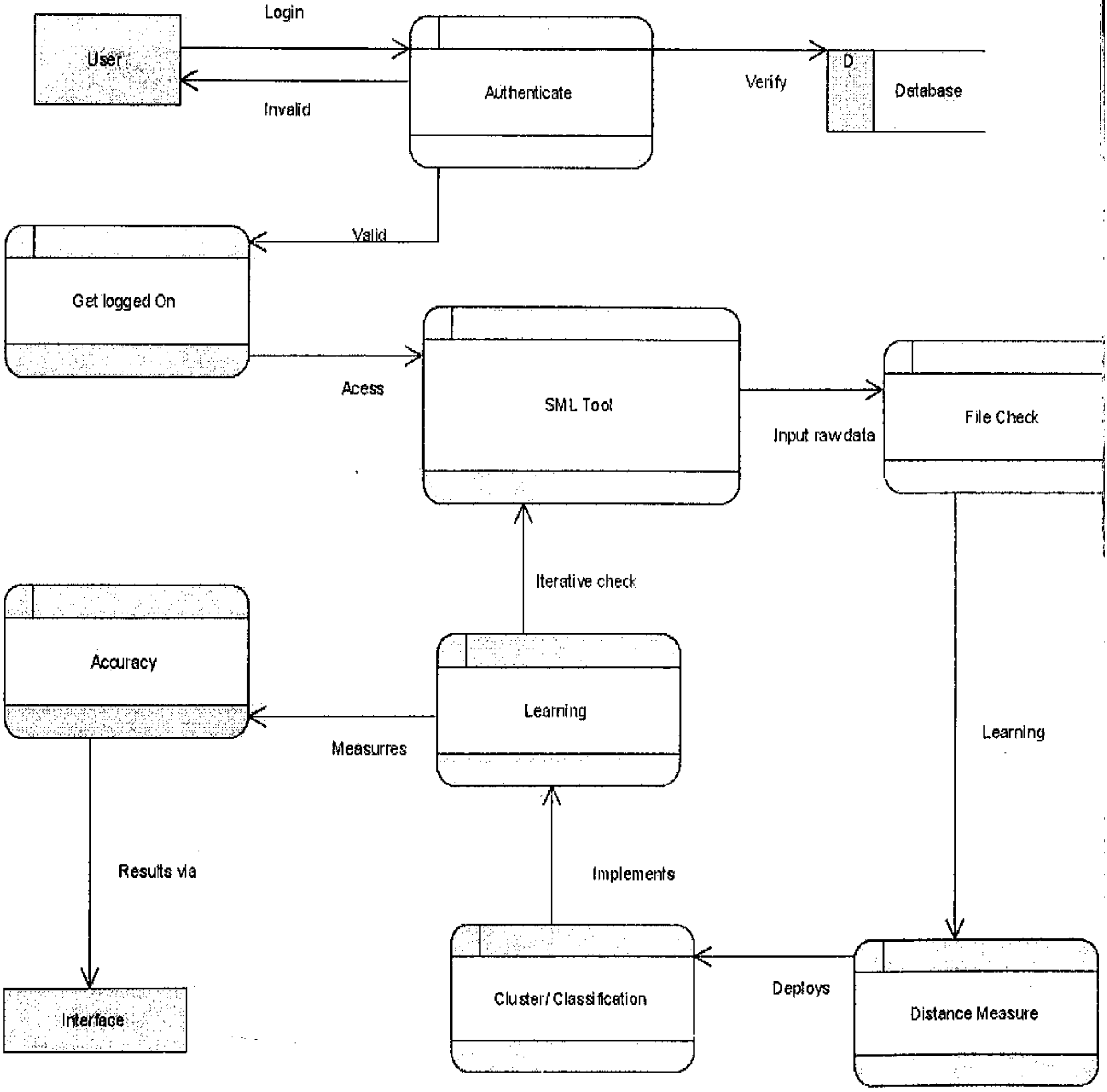


Fig5.2
Level 1 DFD



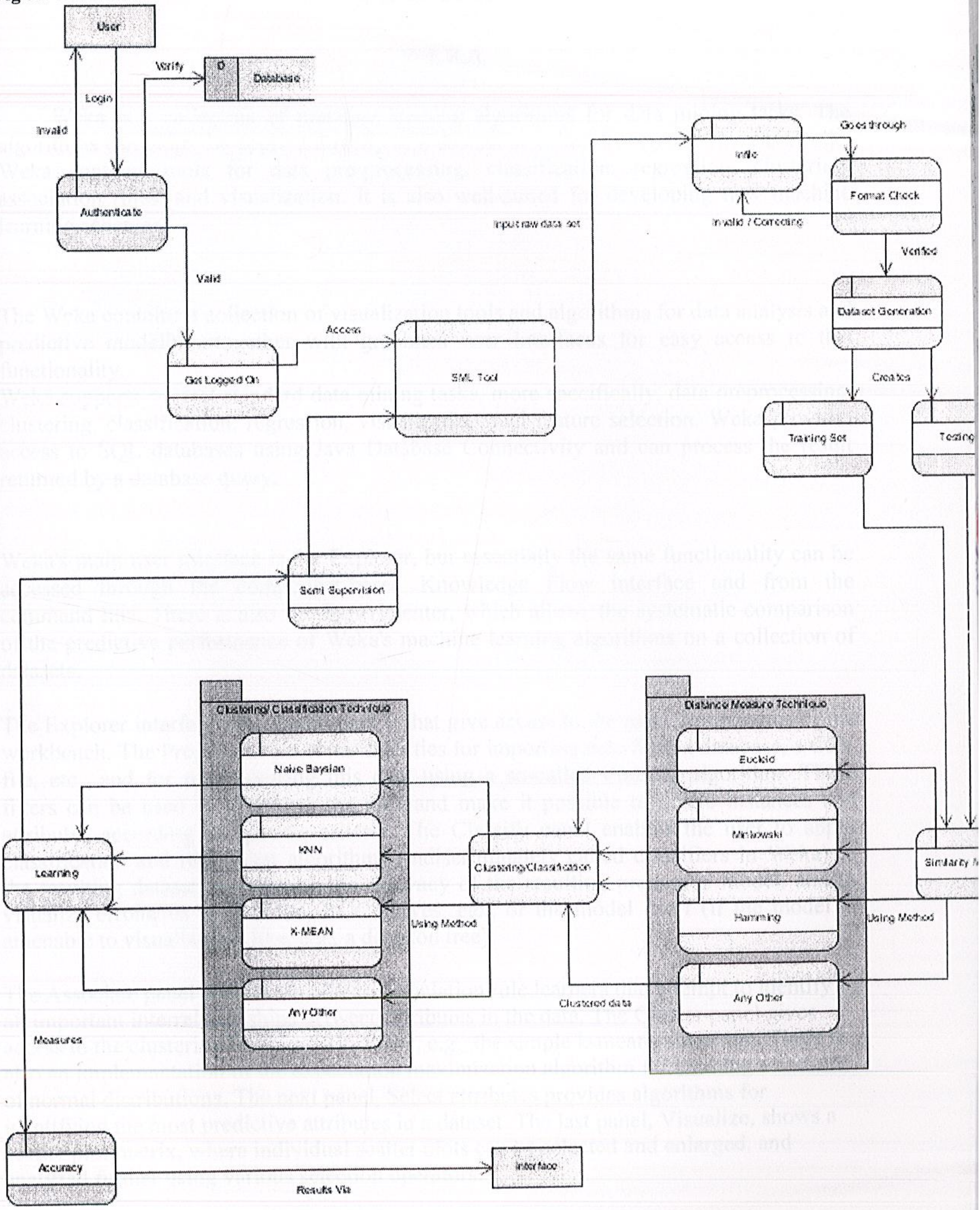
Level 1 DFD

Fig 5.3
Level 2 DFD



Level 2 DFD

Fig 5.4 Level 3 DFD



CHAPTER -6

WEKA

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

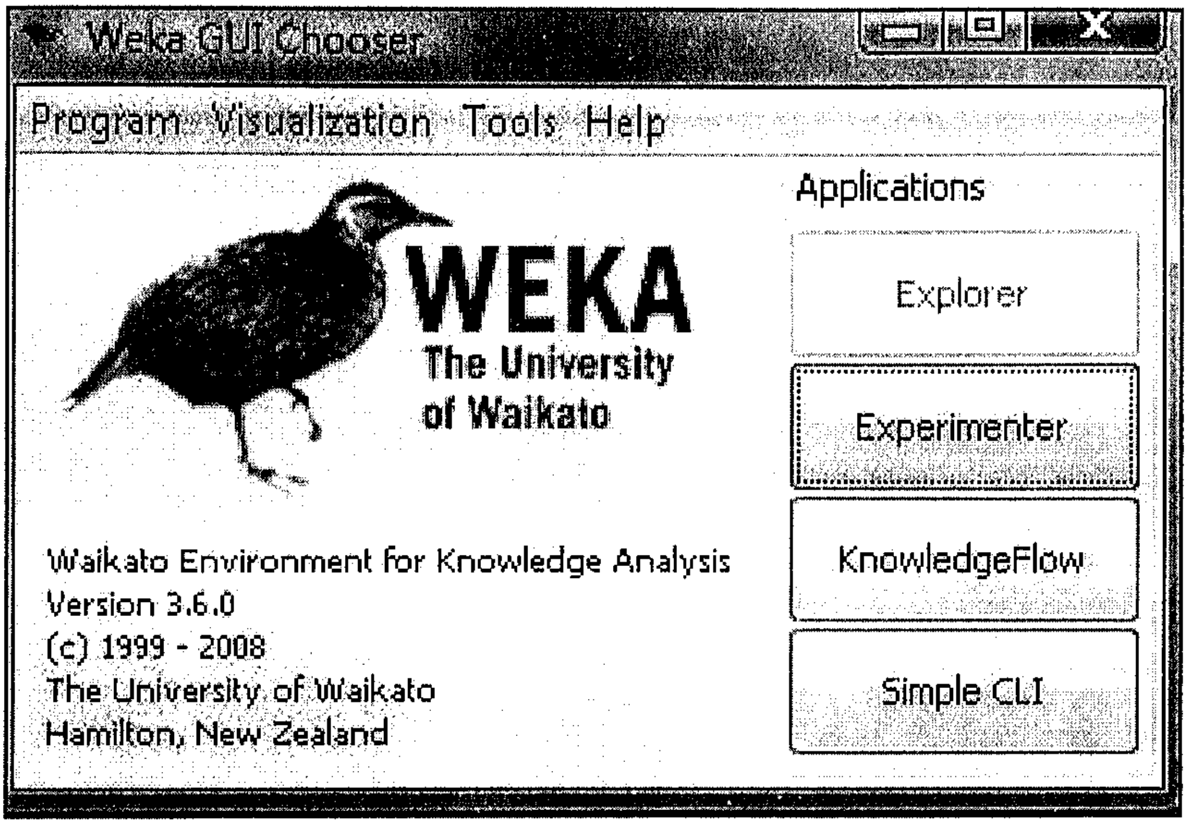
The Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.
Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query.

Weka's main user interface is the Explorer, but essentially the same functionality can be accessed through the component-based Knowledge Flow interface and from the command line. There is also the Experimenter, which allows the systematic comparison of the predictive performance of Weka's machine learning algorithms on a collection of datasets.

The Explorer interface has several panels that give access to the main components of the workbench. The Preprocess panel has facilities for importing data from a database, a CSV file, etc., and for preprocessing this data using a so-called filtering algorithm. These filters can be used to transform the data and make it possible to delete instances and attributes according to specific criteria. The Classify panel enables the user to apply classification and regression algorithms (indiscriminately called classifiers in Weka) to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, ROC curves, etc., or the model itself (if the model is amenable to visualization like, e.g., a decision tree).

The Associate panel provides access to association rule learners that attempt to identify all important interrelationships between attributes in the data. The Cluster panel gives access to the clustering techniques in Weka, e.g., the simple k-means algorithm. There is also an implementation of the expectation maximization algorithm for learning a mixture of normal distributions. The next panel, Select attributes provides algorithms for identifying the most predictive attributes in a dataset. The last panel, Visualize, shows a scatter plot matrix, where individual scatter plots can be selected and enlarged, and analyzed further using various selection operators.

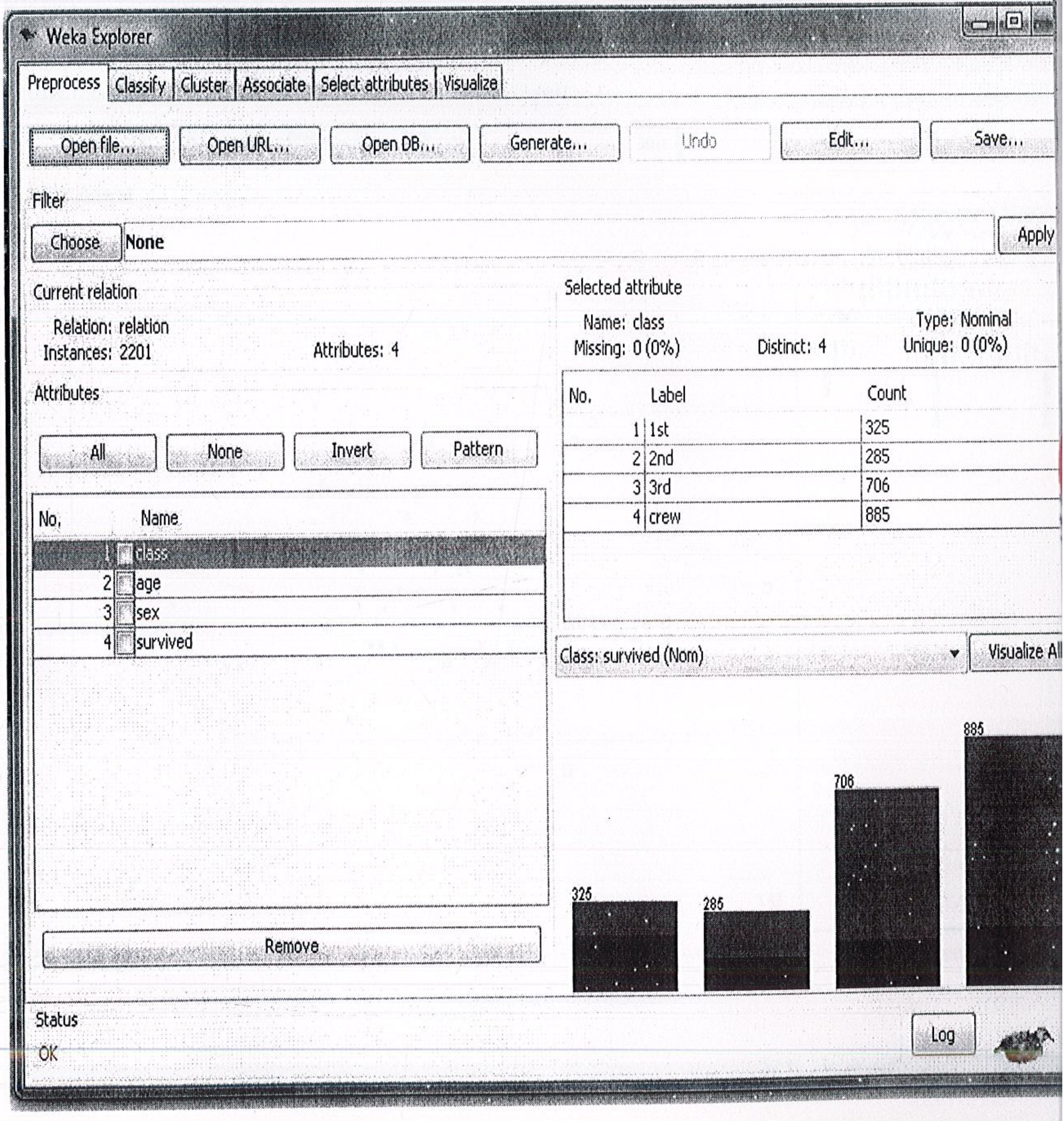
GUI Chooser:



Waikato Environment for Knowledge Analysis
Version 3.6.0
(c) 1999 - 2008
The University of Waikato
Hamilton, New Zealand

- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Significance-testing built in!
- Data sources, classifiers, etc. can be connected graphically
- Data "flows" through components: e.g.,
- "data source" -> "filter" -> "classifier" -> "evaluator"
- Layouts can be saved and loaded again later.

WEKA Explorer



Filter

None

Current relation

Relation: relation
Instances: 2201 Attributes: 4

Attributes

No.	Name
1	□ class
2	□ age
3	□ sex
4	□ survived

Selected attribute

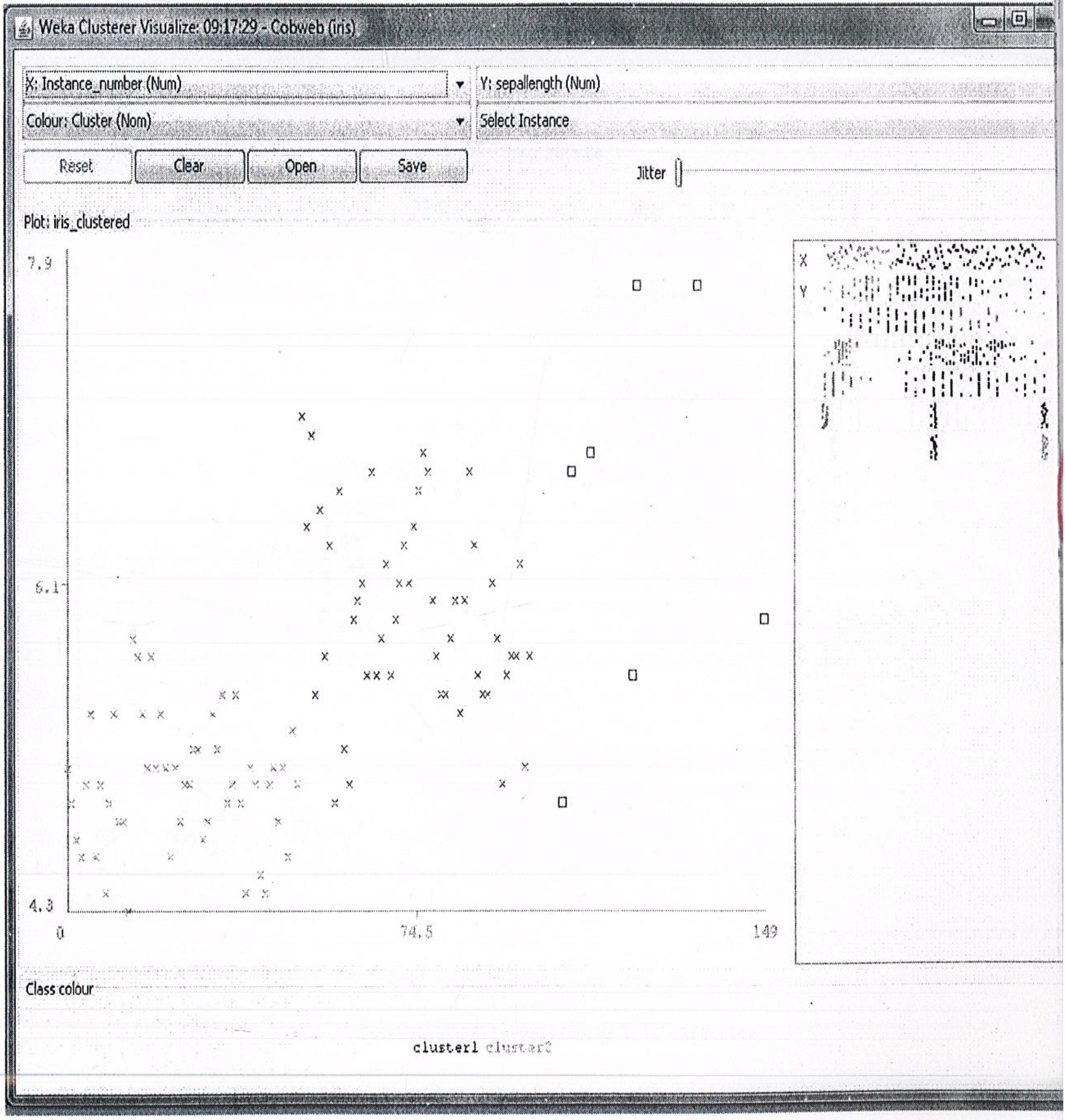
Name: class Type: Nominal
Missing: 0 (0%) Distinct: 4 Unique: 0 (0%)

No.	Label	Count
1	1st	325
2	2nd	285
3	3rd	706
4	crew	885

Class: survived (Nom)

Status
OK

Clustering



Weka Clusterer Visualize: 09:17:29 - Cobweb (iris)

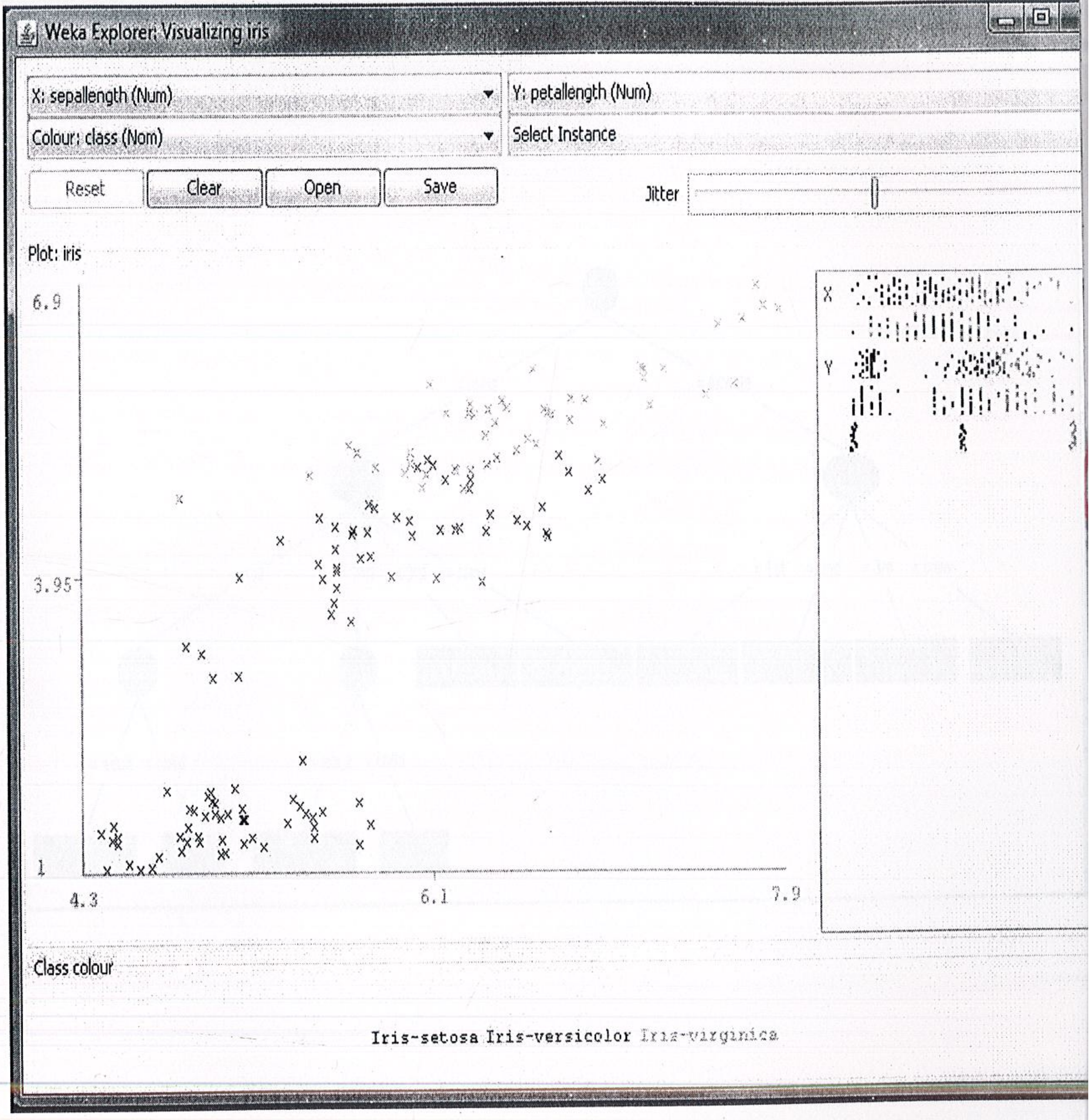
X: Instance_number (Num)

Y: sepallength (Num)

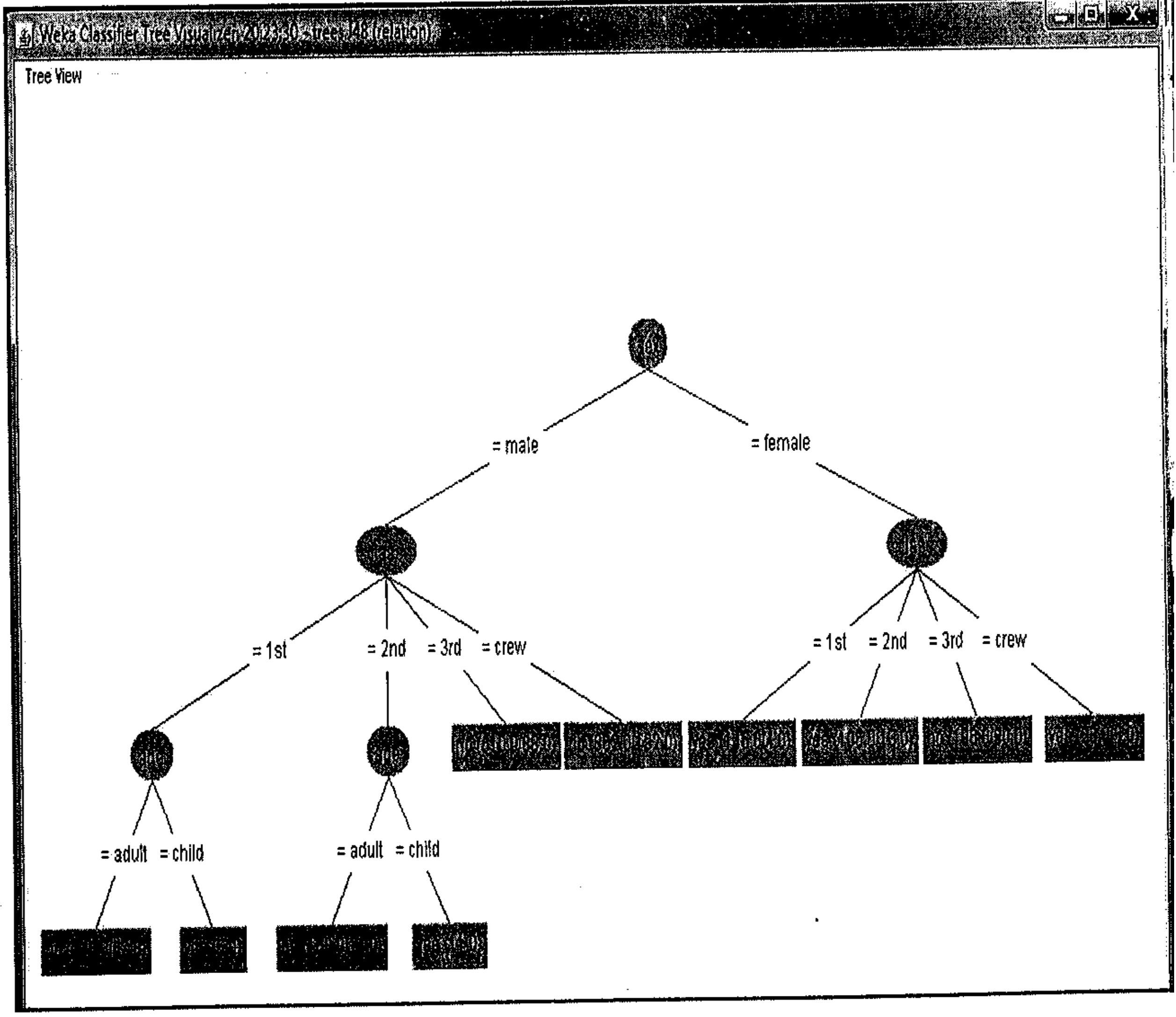
Colour: Cluster (Nom)

Plot: iris_clustered

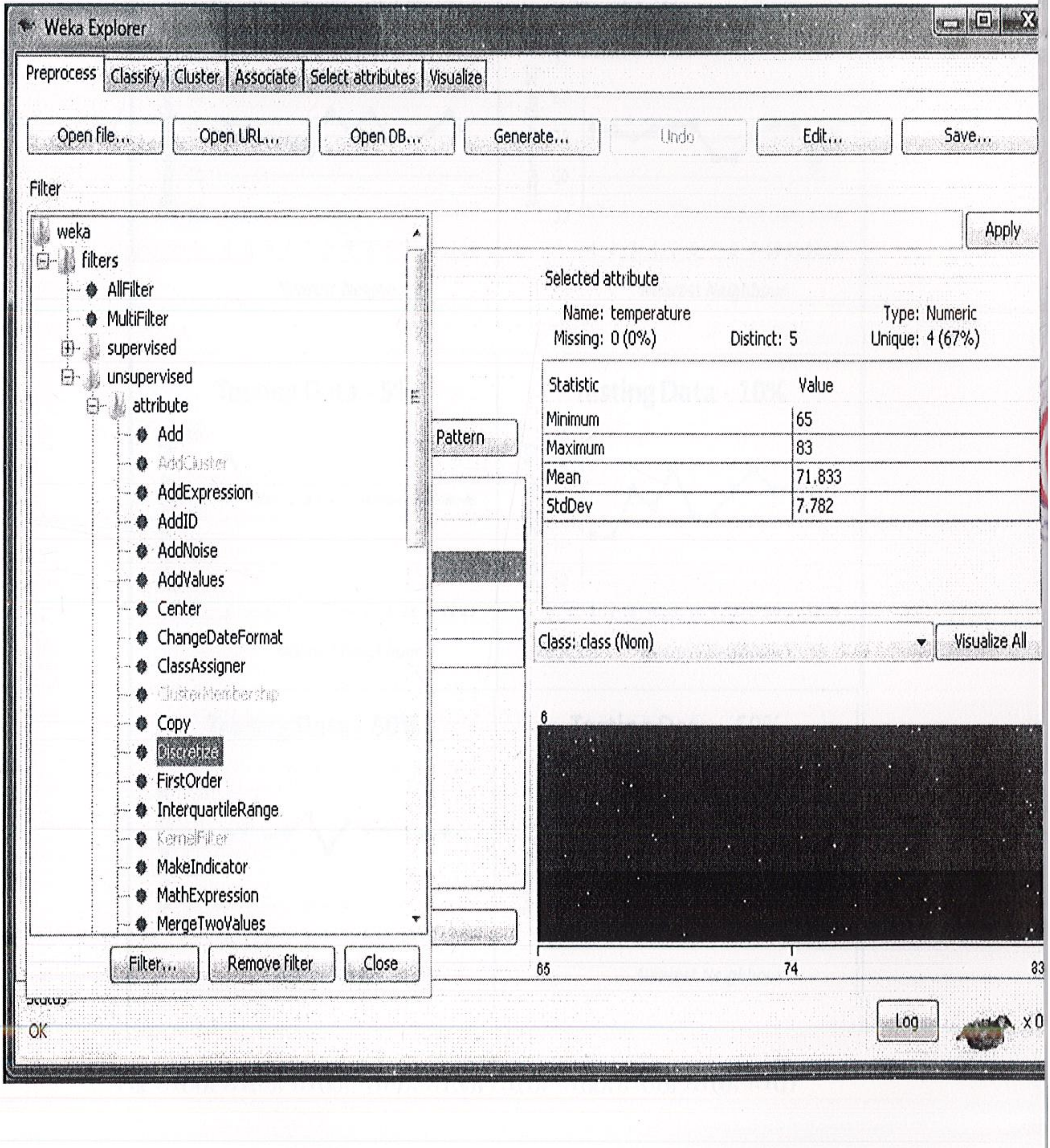
Visualization:



WEKA Tree Visualizer:



WEKA Discretion



Selected attribute

Name: temperature Type: Numeric

Missing: 0 (0%) Distinct: 5 Unique: 4 (67%)

Statistic	Value
Minimum	65
Maximum	83
Mean	71.833
StdDev	7.782

Class: class (Nom)

Status

OK

7.0 Test Output for KNN

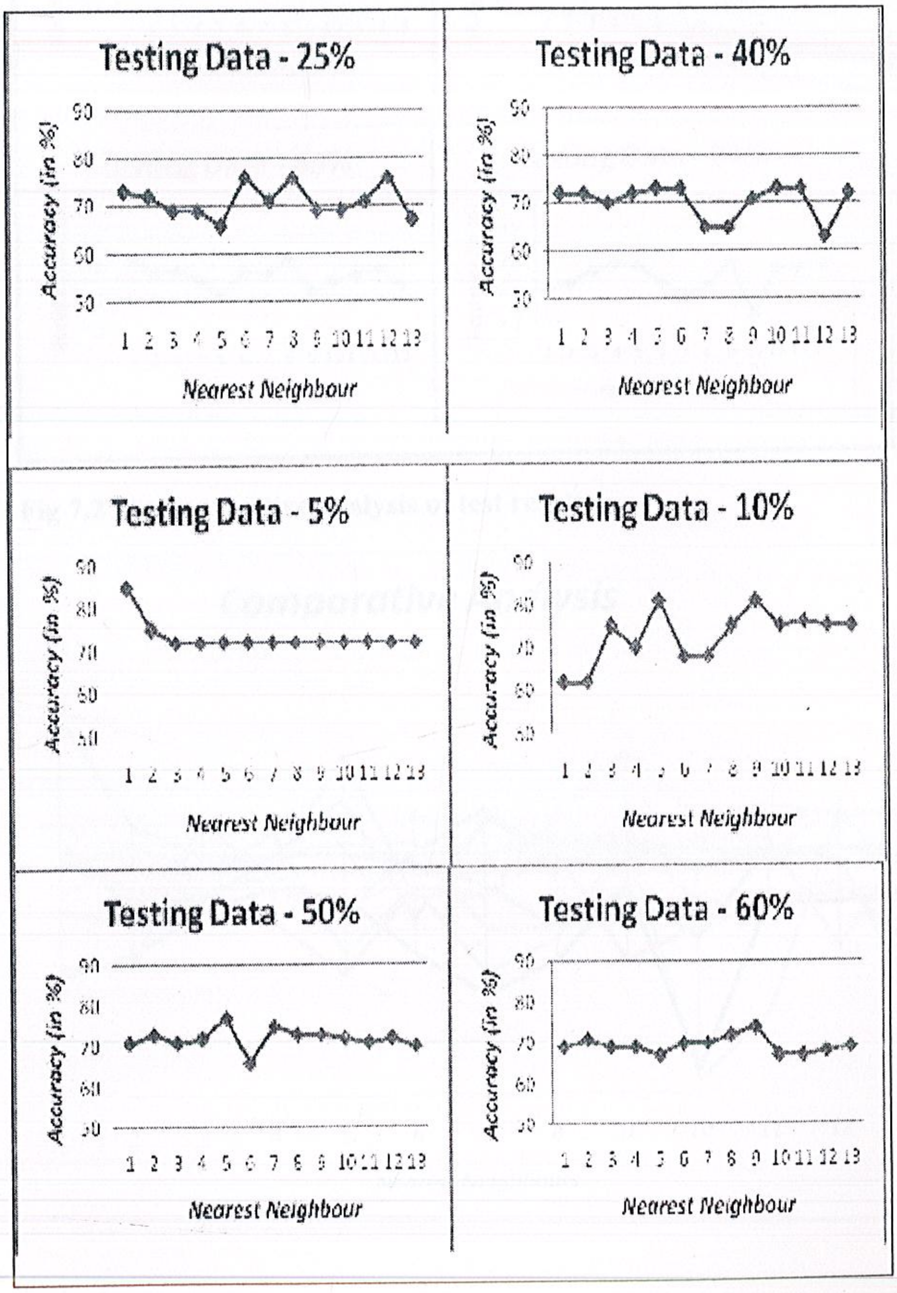


Fig 7.1(a), 7.1(b), 7.1(c), 7.1(d), 7.1(e), 7.1(f), 7.1(g),7.1(h), 7.1(i), 7.1(j)

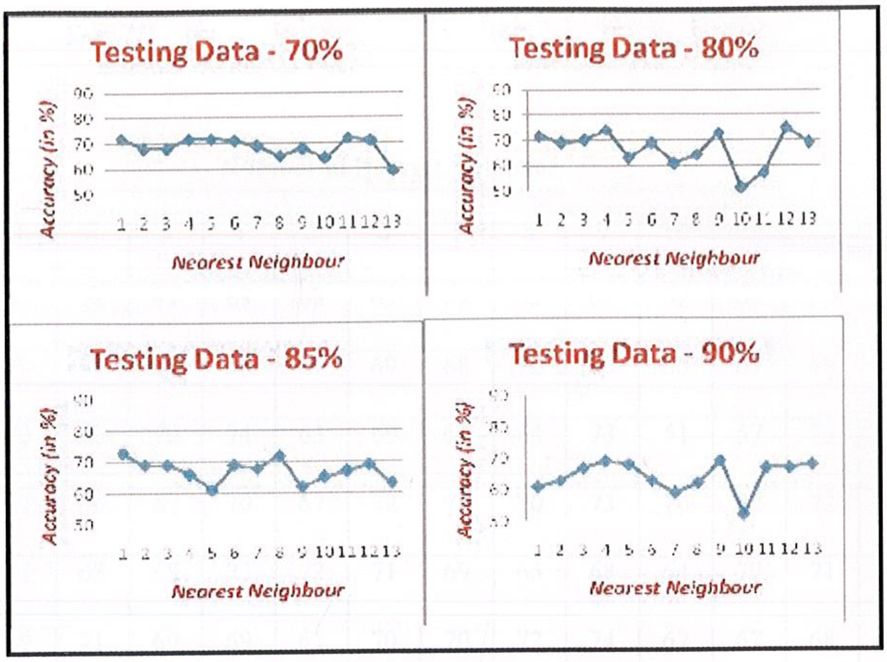
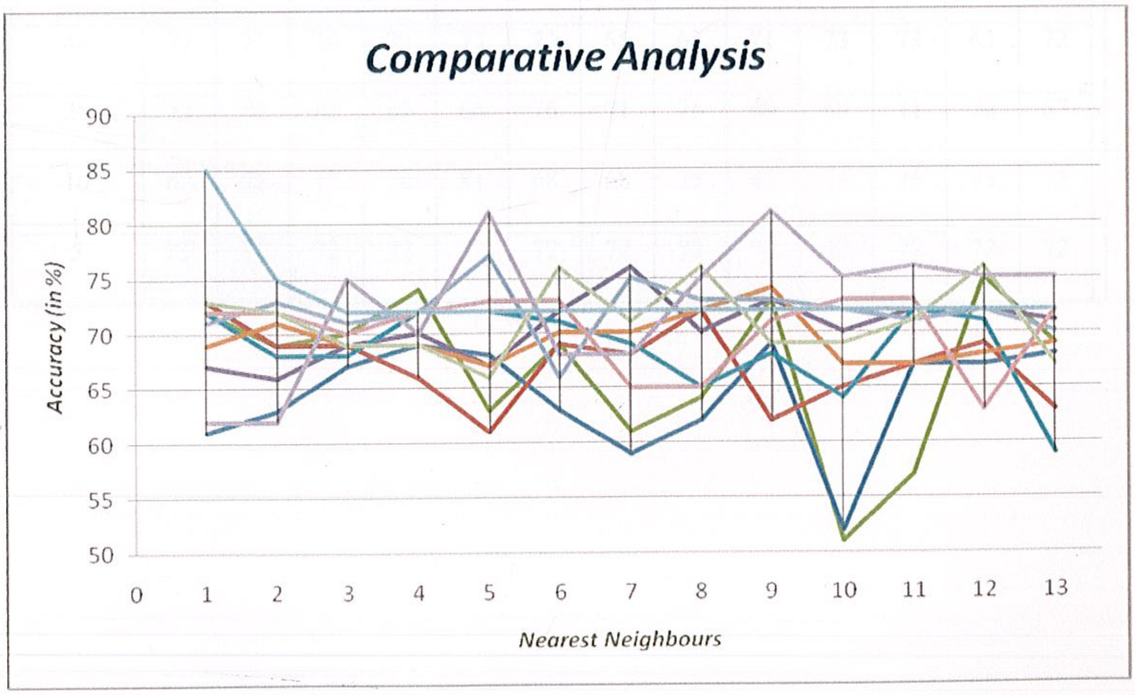


Fig 7.2 :A comparative analysis of test results

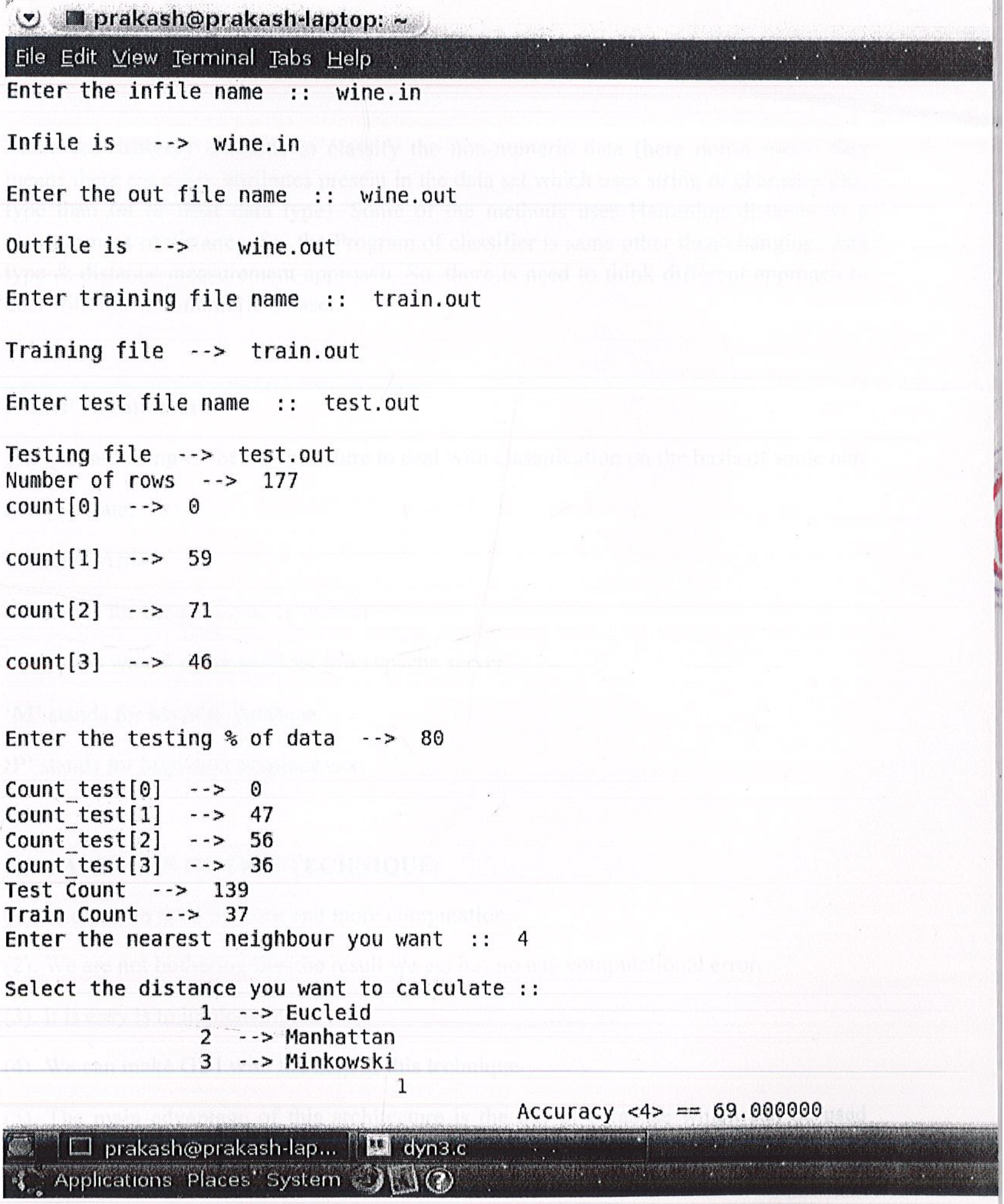
Taining Data (in %)



Number of Nearest Neighbor

	1	2	3	4	5	6	7	8	9	10	11	12	13
90	61	63	67	69	68	63	59	62	69	52	67	67	68
85	73	69	69	66	61	69	68	72	62	65	67	69	63
80	72	69	70	74	63	69	61	64	73	51	57	75	69
75	67	66	69	70	67	72	76	70	73	70	72	72	71
70	72	68	68	72	72	71	69	65	68	64	72	71	59
60	69	71	69	69	67	70	70	72	74	67	67	68	69
50	71	73	71	72	77	66	75	73	73	72	71	72	70
40	72	72	70	72	73	73	65	65	71	73	73	63	72
25	73	72	69	69	66	76	71	76	69	69	71	76	67
10	62	62	75	70	81	68	68	75	81	75	76	75	75
5	85	75	72	72	72	72	72	72	72	72	72	72	72

Fig 7.3 **Screenshot**



prakash@prakash-laptop: ~
File Edit View Terminal Tabs Help
prakash@prakash-lap...
dyn3.c
Applications Places System

```
Enter the infile name  ::  wine.in

Infile is    -->  wine.in

Enter the out file name  ::  wine.out

Outfile is  -->   wine.out

Enter training file name  ::  train.out

Training file  -->  train.out

Enter test file name  ::  test.out

Testing file  -->  test.out
Number of rows  -->  177
count[0]  -->  0

count[1]  -->  59

count[2]  -->  71

count[3]  -->  46


Enter the testing % of data  -->  80

Count_test[0]  -->  0
Count_test[1]  -->  47
Count_test[2]  -->  56
Count_test[3]  -->  36
Test Count  -->  139
Train Count  -->  37
Enter the nearest neighbour you want  ::  4

Select the distance you want to calculate ::
                1  --> Eucleid
                2  --> Manhattan
                3  --> Minkowski
                                1
                                            Accuracy <4> == 69.000000
```

CHAPTER 8

HOW TO DEAL WITH NON-NUMERIC DATA TO CLASSIFY IT?

There are different methods to classify the non-numeric data (here non-numeric data means there are many attributes present in the data set which uses string or character data type than int or float data type). Some of the methods uses Hamming distance as a measurement of distance. So, the Program of classifier is same other then changing data type & distance measurement approach. So, there is need to think different approach to deal with this non-numeric dataset.

LAMP Architecture

Here we are using LAMP architecture to deal with classification on the basis of some non numeric data.

What is LAMP?

'**L**' stands for Linux operating system.

'**A**' means we are using services from apache server.

'**M**' stands for MySQL database.

'**P**' stands for hypertext preprocessor.

ADVANTAGES OF THIS TECHNIQUE:

(1). There is no need of more and more computation.

(2). We are not bothering that the result we get has no any computational error.

(3). It is easy is to implement.

(4). We can make GUI with the help of this technique.

(5). The main advantage of this architecture is the use of database which can be used again & again for further computation.

LAMP is an acronym for a solution stack of free, open source software, originally coined from the first letters of Linux (operating system), Apache HTTP Server, MySQL, and PHP, principal components to build a viable general purpose web server.

The precise combination of software included in a LAMP package may vary, especially with respect to the web scripting software, as PHP may be replaced by Perl or Python. Similar terms exist for essentially the same software suite (AMP) running on other operating systems, such as MS Windows (WAMP), Mac OS (MAMP), Solaris (SAMP), or OpenBSD (OpAMP).

When used in combination they represent a solution stack of technologies that support application servers.

The LAMP stack is widely used because it offers a great number of advantages for developers:

- Easy to code: Novices can build something and get it up and running very quickly with PHP and MySQL.
- Easy to deploy: Since PHP is a standard Apache module, it's easy to deploy a PHP app. Once you've got MySQL running, simply upload your .php files.
- Develop locally: It's easy to set up LAMP on your laptop, build your app locally, then deploy on the Web.
- Cheap and ubiquitous hosting: Even the cheapest Web hosts options allow you to run PHP and MySQL.

8.1 LINUX APPROACH:

Here we are using the simple linux commands to deal with our original data. Some of few are:

- grep command,
- Cut command,
- Paste command,

e.g. if we are to deal with adult.data.txt data set we are using above command as such are using above command as such.

Diagrammatical Implementation of Approach

DFD Level 0:

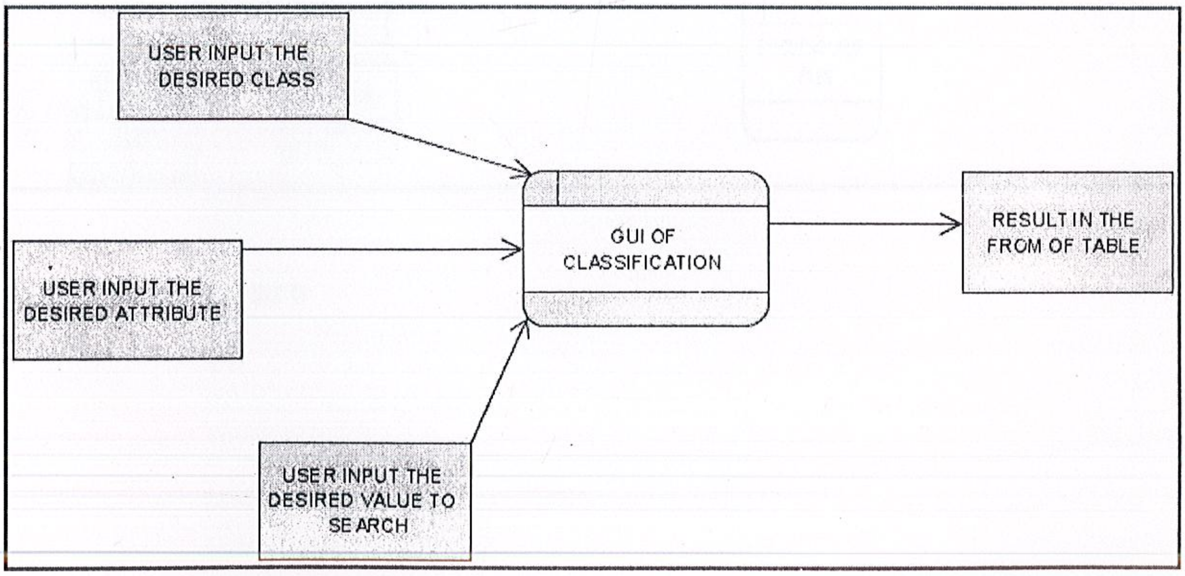

USER INPUT THE DESIRED CLASS
USER INPUT THE DESIRED ATTRIBUTE
USER INPUT THE DESIRED VALUE TO SEARCH
GUI OF CLASSIFICATION
RESULT IN THE FROM OF TABLE

Figure 8.1 : DFD Level 0

DFD Level 1:

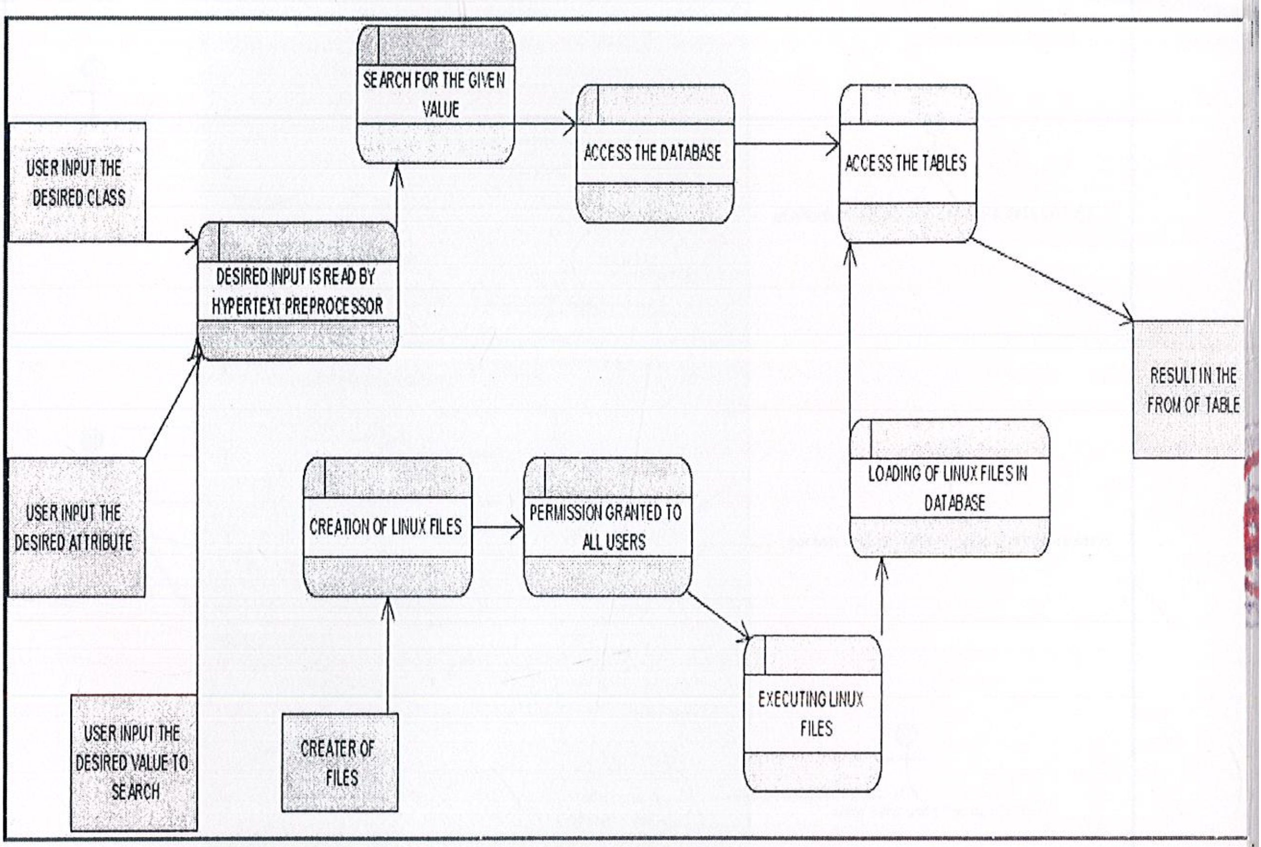

SEARCH FOR THE GIVEN VALUE
ACCESS THE DATABASE
ACCESS THE TABLES
USER INPUT THE DESIRED CLASS
DESIRED INPUT IS READ BY HYPERTEXT PREPROCESSOR
RESULT IN THE FROM OF TABLE
USER INPUT THE DESIRED ATTRIBUTE
CREATION OF LINUX FILES
PERMISSION GRANTED TO ALL USERS
LOADING OF LINUX FILES IN DATABASE
EXECUTING LINUX FILES
USER INPUT THE DESIRED VALUE TO SEARCH
CREATER OF FILES

Figure 8.2 : Level 1 DFD

Use Case Diagram :

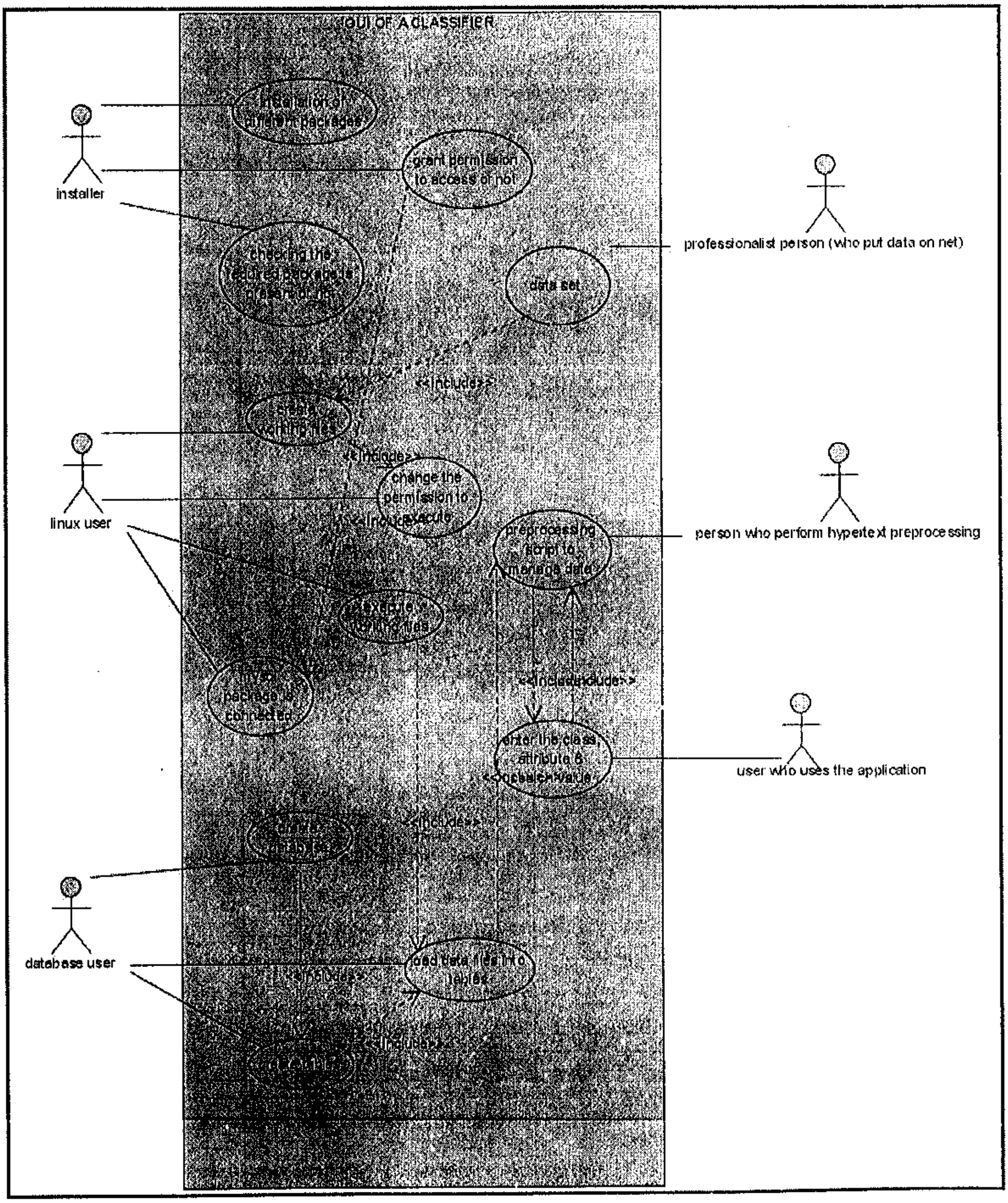

GUI OF A CLASSIFIER
installer
installation of different packages
grant permission to access or not
checking the required package is present or not
data set
professionalist person (who put data on net)
<<include>>
create working files
<<include>>
linux user
change the permission to execute
<<include>>
preprocessing script to manage data
person who perform hypertext preprocessing
execute working files
mysql package is connected
<<includeinclude>>
enter the class, attribute & search value
user who uses the application
<<include>>
database user
load data files into tables
<<include>>
<<include>>

Figure 8.3 : USE CASE DIAGRAM

Sequence Diagram:

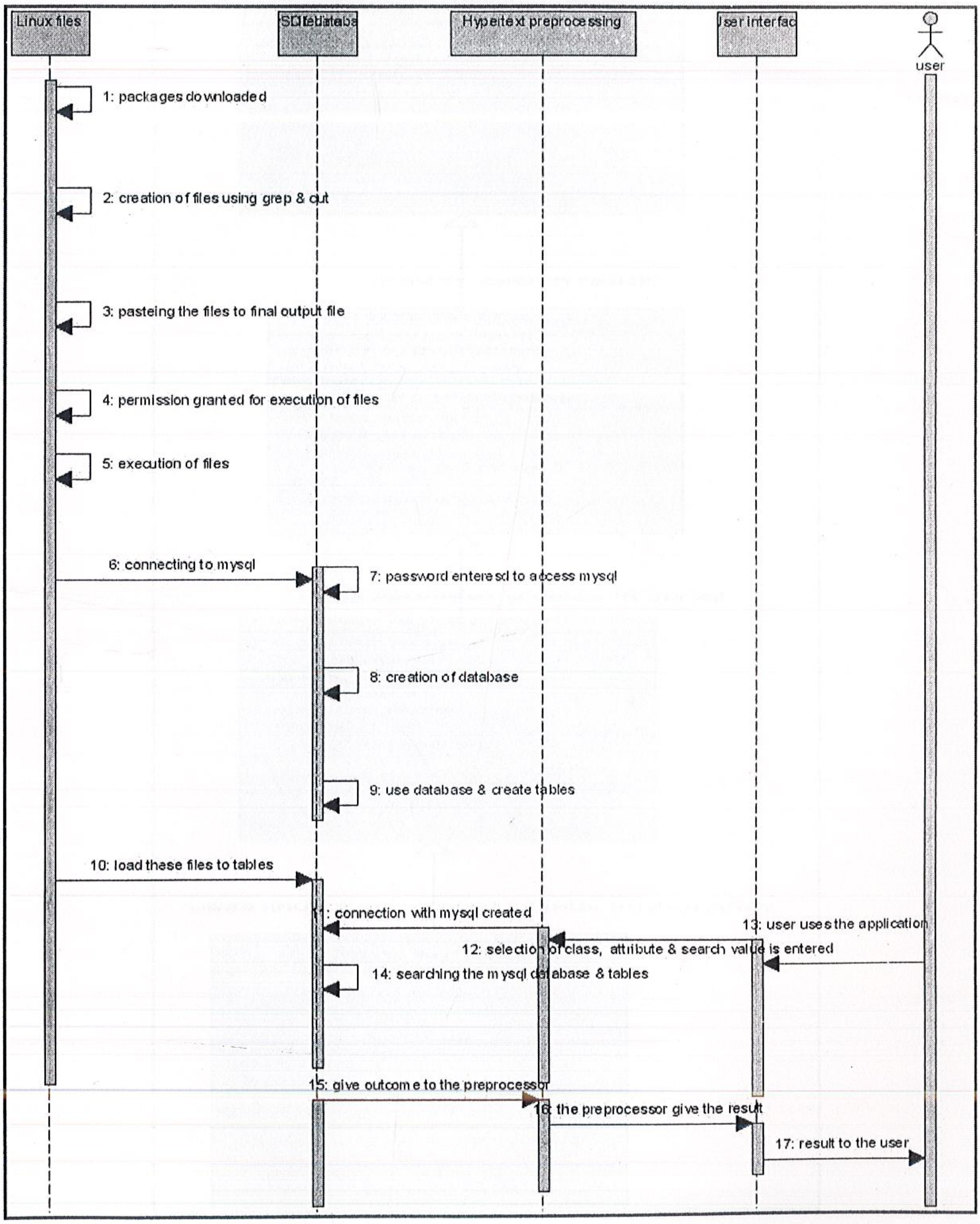

Linux files
SQlfedatabas
Hypeitext preprocessing
ser interfac
user
1: packages downloaded
2: creation of files using grep & cut
3: pasteing the files to final output file
4: permission granted for execution of files
5: execution of files
6: connecting to mysql
7: password enteresd to access mysql
8: creation of database
9: use database & create tables
10: load these files to tables
11: connection with mysql created
12: selection of class, attribute & search value is entered
13: user uses the application
14: searching the mysql database & tables
15: give outcome to the preprocessor
16: the preprocessor give the result
17: result to the user

Figure 8.4 : **Sequence Diagram**

Class Diagram:

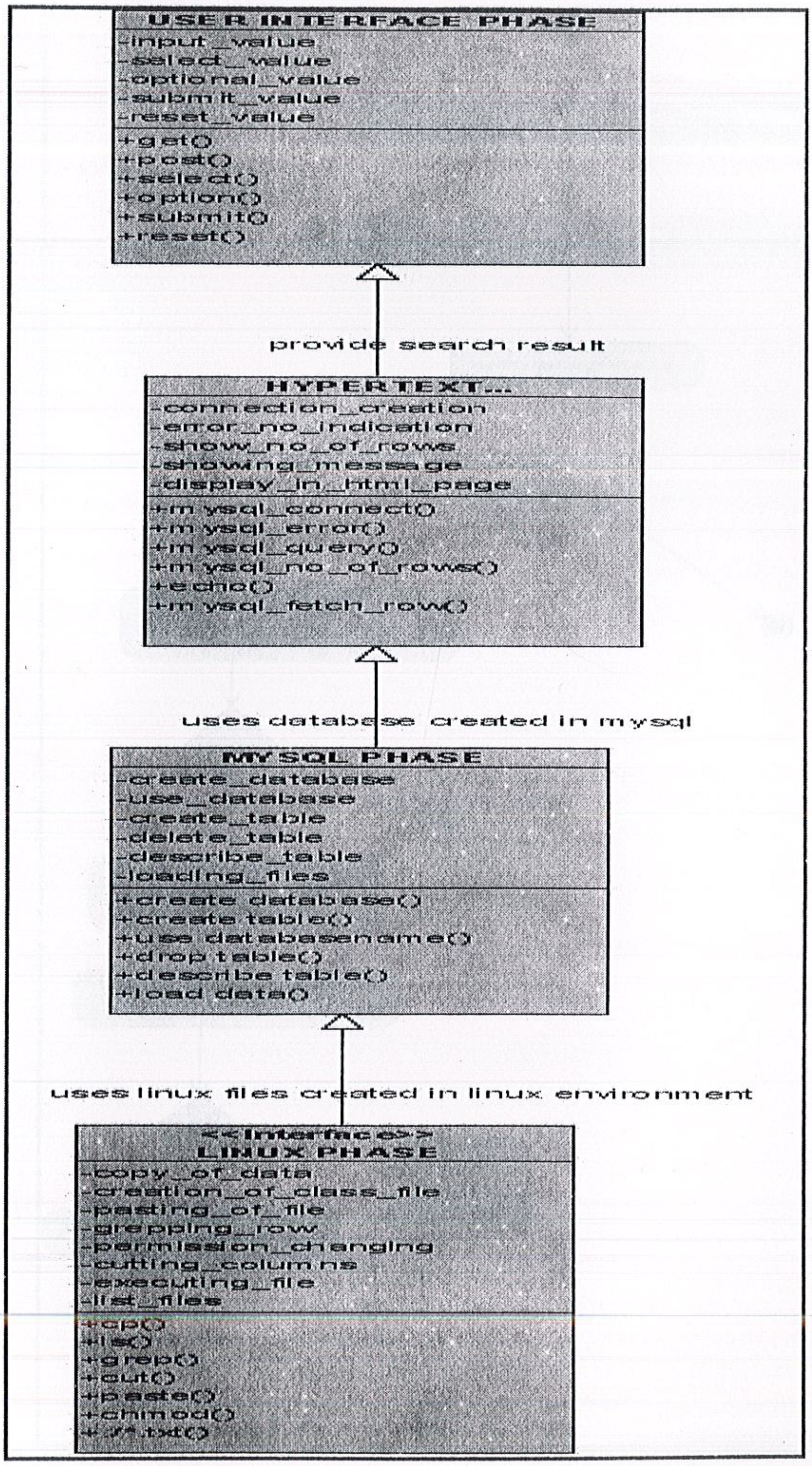


Figure 8.5 : Class Diagram

Activity Diagram:

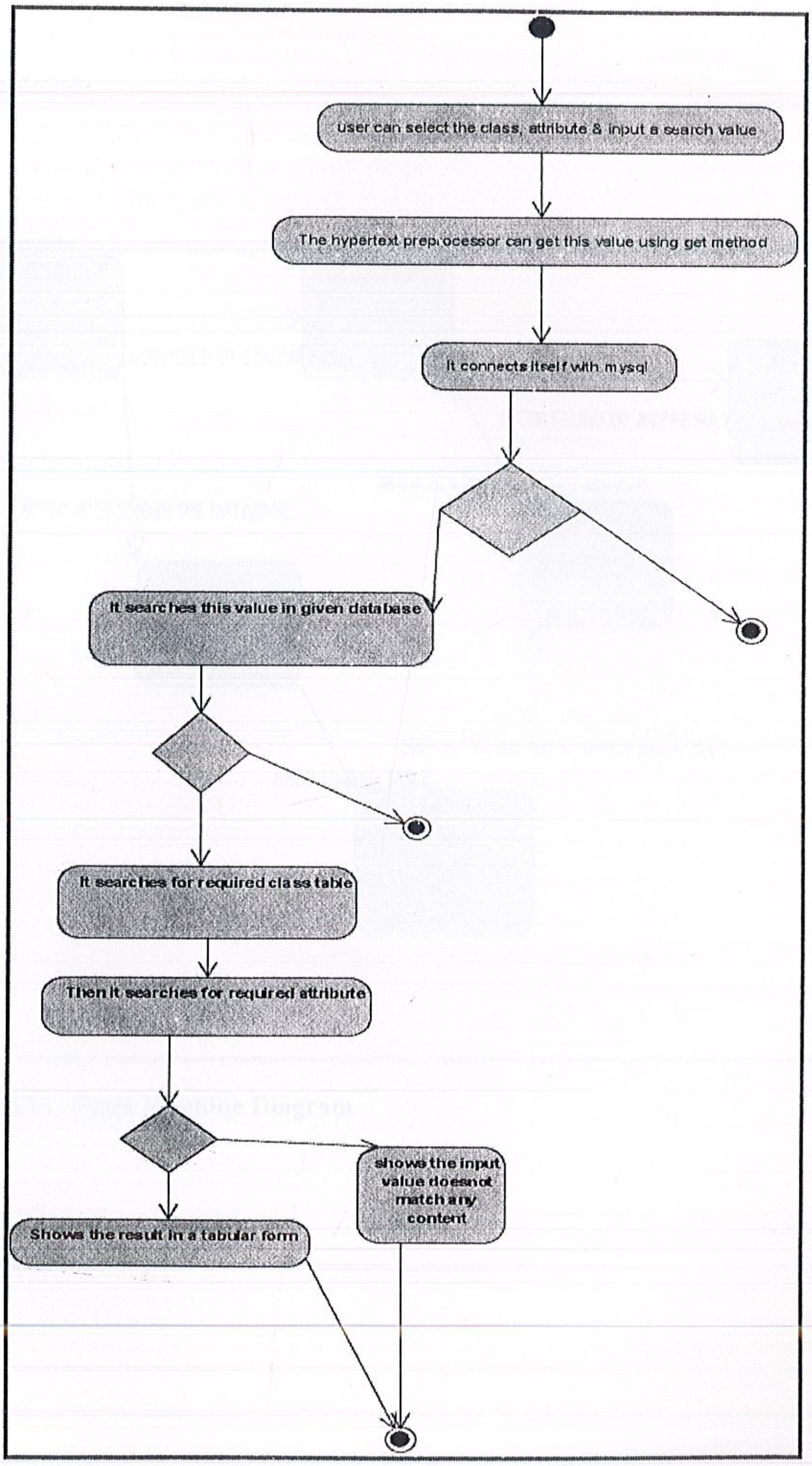

user can select the class, attribute & input a search value
The hypertext preprocessor can get this value using get method
It connects itself with mysql
It searches this value in given database
It searches for required class table
Then it searches for required attribute
Shows the result in a tabular form
shows the input value doesnot match any content

Figure 8.6 : Activity Diagram

State Machine Diagram:

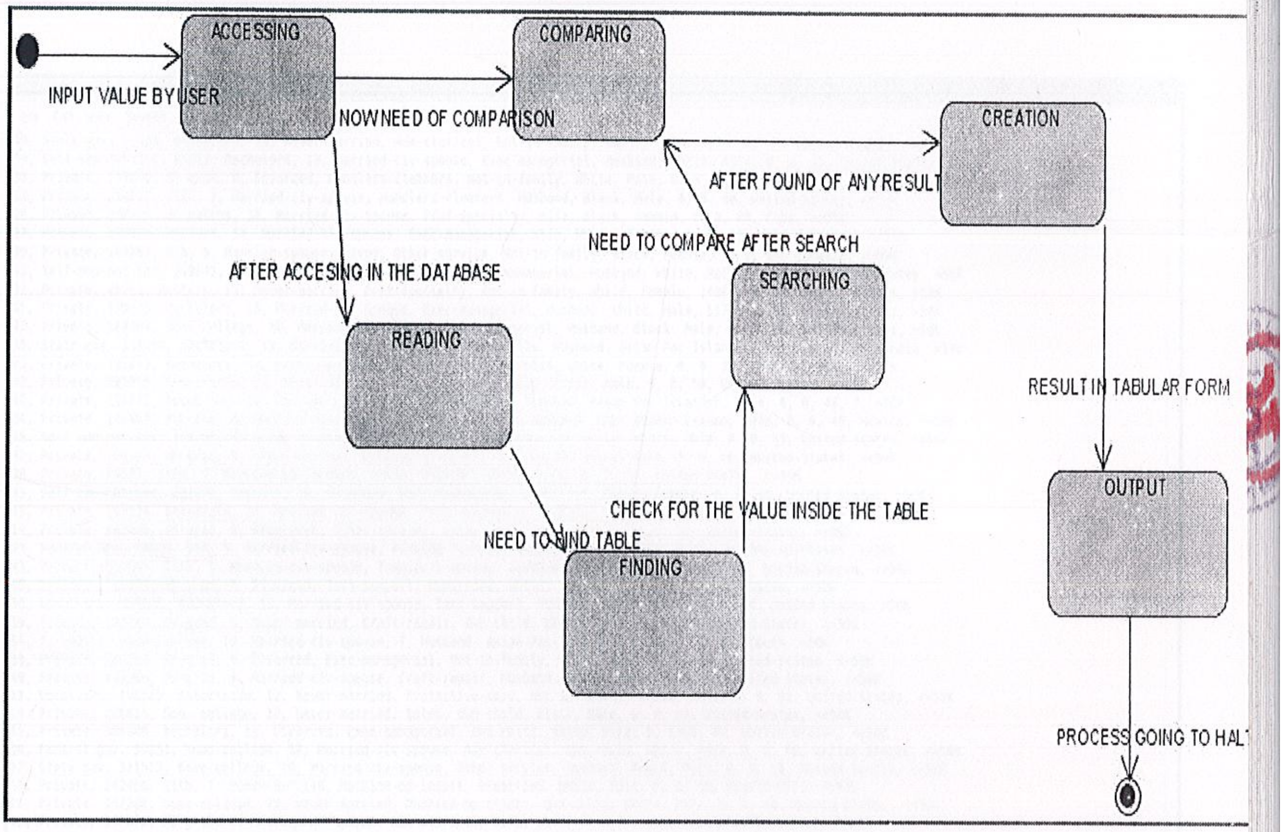
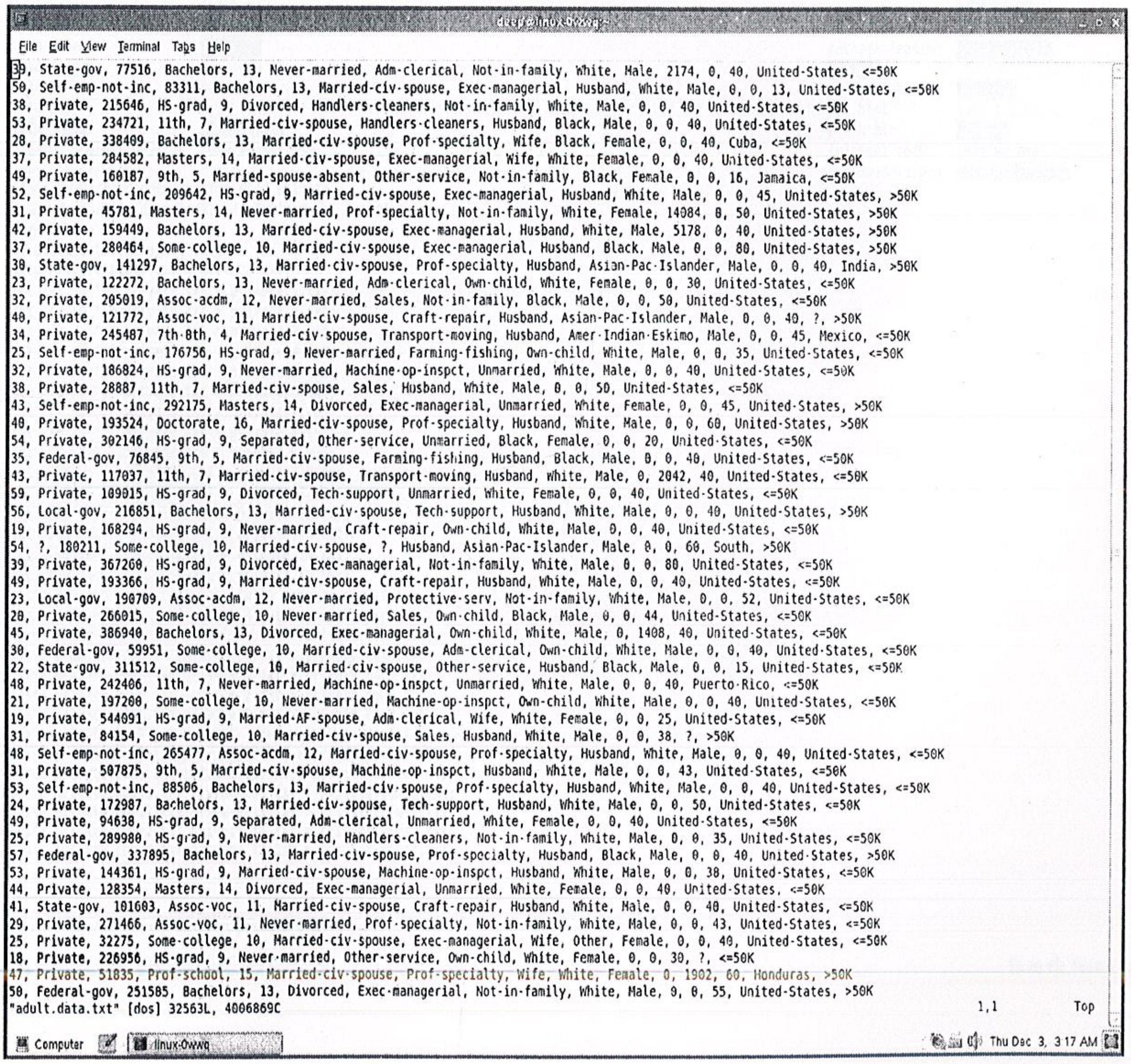


Figure 8.7 : State Machine Diagram

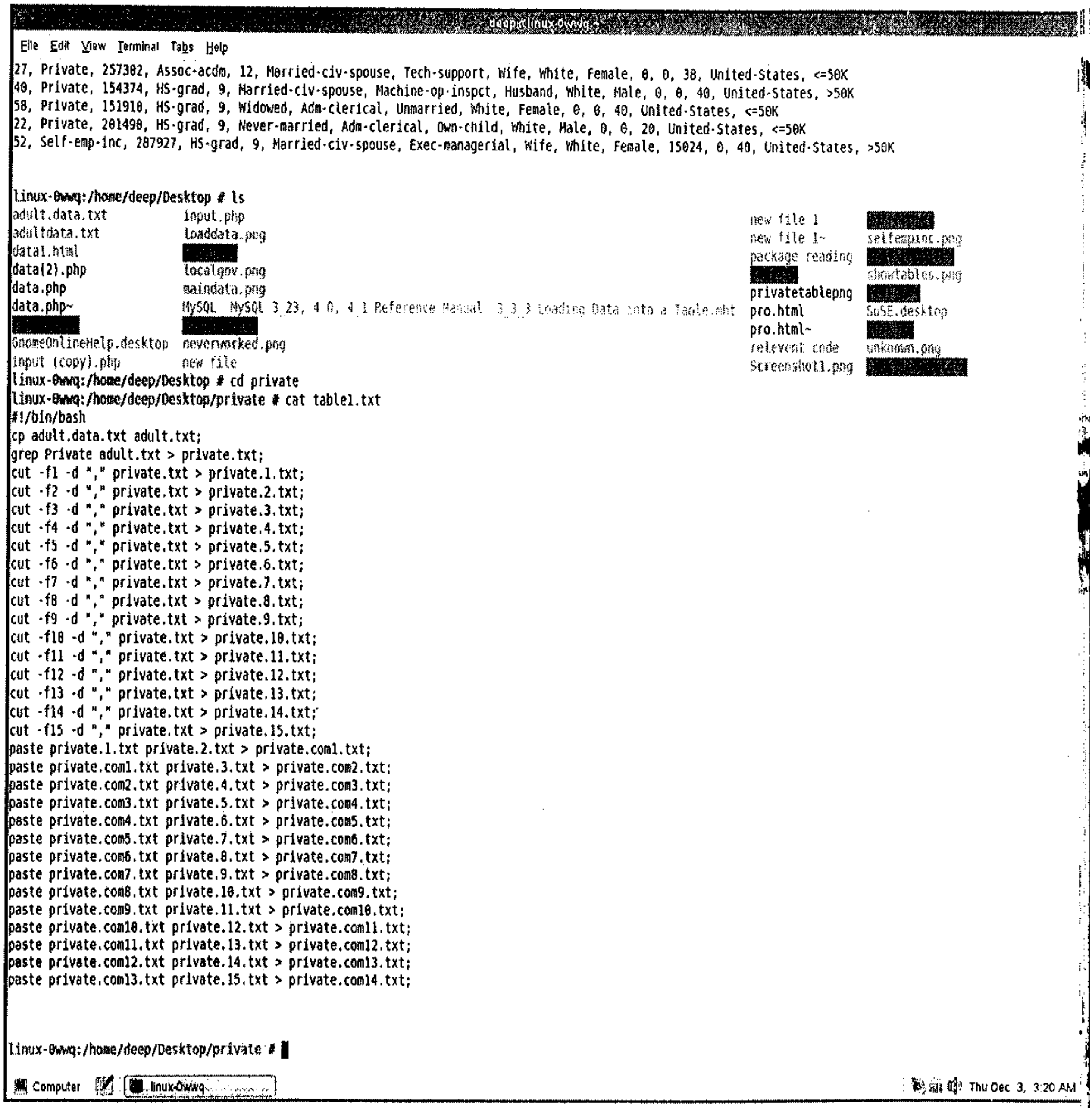
Screenshots for Current Approach (DEALING WITH NON-NUMERIC DATA TO CLASSIFY):

Main Data:



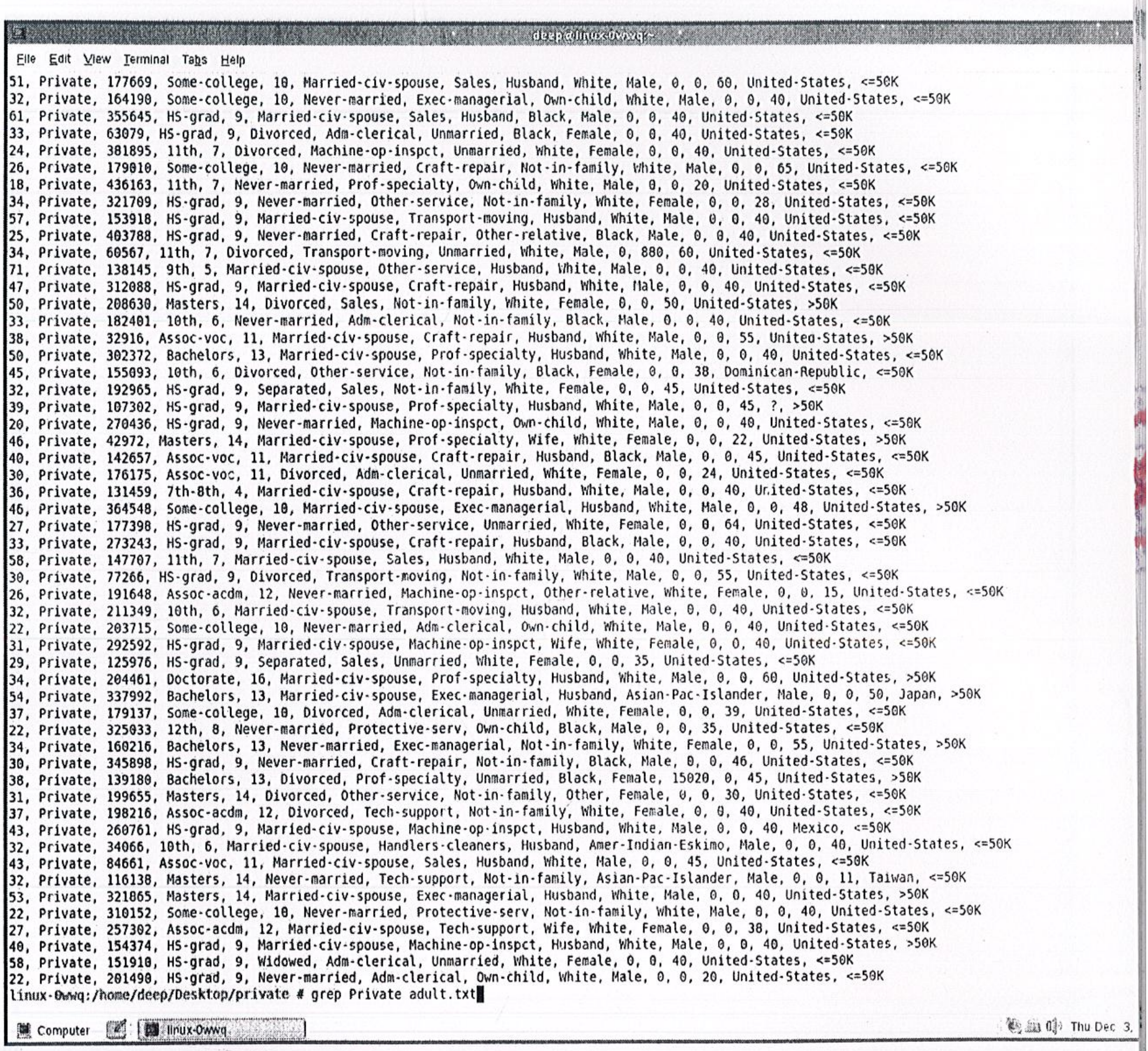
```
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
37, Private, 284582, Masters, 14, Married-civ-spouse, Exec-managerial, Wife, White, Female, 0, 0, 40, United-States, <=50K
49, Private, 160187, 9th, 5, Married-spouse-absent, Other-service, Not-in-family, Black, Female, 0, 0, 16, Jamaica, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 45, United-States, >50K
31, Private, 45781, Masters, 14, Never-married, Prof-specialty, Not-in-family, White, Female, 14084, 0, 50, United-States, >50K
42, Private, 159449, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 5178, 0, 40, United-States, >50K
37, Private, 280464, Some-college, 10, Married-civ-spouse, Exec-managerial, Husband, Black, Male, 0, 0, 80, United-States, >50K
30, State-gov, 141297, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, Asian-Pac-Islander, Male, 0, 0, 40, India, >50K
23, Private, 122272, Bachelors, 13, Never-married, Adm-clerical, Own-child, White, Female, 0, 0, 30, United-States, <=50K
32, Private, 205019, Assoc-acdm, 12, Never-married, Sales, Not-in-family, Black, Male, 0, 0, 50, United-States, <=50K
40, Private, 121772, Assoc-voc, 11, Married-civ-spouse, Craft-repair, Husband, Asian-Pac-Islander, Male, 0, 0, 40, ?, >50K
34, Private, 245487, 7th-8th, 4, Married-civ-spouse, Transport-moving, Husband, Amer-Indian-Eskimo, Male, 0, 0, 45, Mexico, <=50K
25, Self-emp-not-inc, 176756, HS-grad, 9, Never-married, Farming-fishing, Own-child, White, Male, 0, 0, 35, United-States, <=50K
32, Private, 186824, HS-grad, 9, Never-married, Machine-op-inspct, Unmarried, White, Male, 0, 0, 40, United-States, <=50K
38, Private, 28887, 11th, 7, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 50, United-States, <=50K
43, Self-emp-not-inc, 292175, Masters, 14, Divorced, Exec-managerial, Unmarried, White, Female, 0, 0, 45, United-States, >50K
40, Private, 193524, Doctorate, 16, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 60, United-States, >50K
54, Private, 302146, HS-grad, 9, Separated, Other-service, Unmarried, Black, Female, 0, 0, 20, United-States, <=50K
35, Federal-gov, 76845, 9th, 5, Married-civ-spouse, Farming-fishing, Husband, Black, Male, 0, 0, 40, United-States, <=50K
43, Private, 117037, 11th, 7, Married-civ-spouse, Transport-moving, Husband, White, Male, 0, 2042, 40, United-States, <=50K
59, Private, 109015, HS-grad, 9, Divorced, Tech-support, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
56, Local-gov, 216851, Bachelors, 13, Married-civ-spouse, Tech-support, Husband, White, Male, 0, 0, 40, United-States, >50K
19, Private, 168294, HS-grad, 9, Never-married, Craft-repair, Own-child, White, Male, 0, 0, 40, United-States, <=50K
54, ?, 180211, Some-college, 10, Married-civ-spouse, ?, Husband, Asian-Pac-Islander, Male, 0, 0, 60, South, >50K
39, Private, 367260, HS-grad, 9, Divorced, Exec-managerial, Not-in-family, White, Male, 0, 0, 80, United-States, <=50K
49, Private, 193366, HS-grad, 9, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 40, United-States, <=50K
23, Local-gov, 190709, Assoc-acdm, 12, Never-married, Protective-serv, Not-in-family, White, Male, 0, 0, 52, United-States, <=50K
20, Private, 266015, Some-college, 10, Never-married, Sales, Own-child, Black, Male, 0, 0, 44, United-States, <=50K
45, Private, 386940, Bachelors, 13, Divorced, Exec-managerial, Own-child, White, Male, 0, 1408, 40, United-States, <=50K
30, Federal-gov, 59951, Some-college, 10, Married-civ-spouse, Adm-clerical, Own-child, White, Male, 0, 0, 40, United-States, <=50K
22, State-gov, 311512, Some-college, 10, Married-civ-spouse, Other-service, Husband, Black, Male, 0, 0, 15, United-States, <=50K
48, Private, 242406, 11th, 7, Never-married, Machine-op-inspct, Unmarried, White, Male, 0, 0, 40, Puerto-Rico, <=50K
21, Private, 197200, Some-college, 10, Never-married, Machine-op-inspct, Own-child, White, Male, 0, 0, 40, United-States, <=50K
19, Private, 544091, HS-grad, 9, Married-AF-spouse, Adm-clerical, Wife, White, Female, 0, 0, 25, United-States, <=50K
31, Private, 84154, Some-college, 10, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 38, ?, >50K
48, Self-emp-not-inc, 265477, Assoc-acdm, 12, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 40, United-States, <=50K
31, Private, 507875, 9th, 5, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 43, United-States, <=50K
53, Self-emp-not-inc, 88506, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 40, United-States, <=50K
24, Private, 172987, Bachelors, 13, Married-civ-spouse, Tech-support, Husband, White, Male, 0, 0, 50, United-States, <=50K
49, Private, 94638, HS-grad, 9, Separated, Adm-clerical, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
25, Private, 289980, HS-grad, 9, Never-married, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 35, United-States, <=50K
57, Federal-gov, 337895, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, Black, Male, 0, 0, 40, United-States, >50K
53, Private, 144361, HS-grad, 9, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 38, United-States, <=50K
44, Private, 128354, Masters, 14, Divorced, Exec-managerial, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
41, State-gov, 101603, Assoc-voc, 11, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 40, United-States, <=50K
29, Private, 271466, Assoc-voc, 11, Never-married, Prof-specialty, Not-in-family, White, Male, 0, 0, 43, United-States, <=50K
25, Private, 32275, Some-college, 10, Married-civ-spouse, Exec-managerial, Wife, Other, Female, 0, 0, 40, United-States, <=50K
18, Private, 226956, HS-grad, 9, Never-married, Other-service, Own-child, White, Female, 0, 0, 30, ?, <=50K
47, Private, 51835, Prof-school, 15, Married-civ-spouse, Prof-specialty, Wife, White, Female, 0, 1902, 60, Honduras, >50K
50, Federal-gov, 251585, Bachelors, 13, Divorced, Exec-managerial, Not-in-family, White, Male, 9, 0, 55, United-States, >50K
"adult.data.txt" [dos] 32563L, 4006869C                                                            1,1           Top
```

Private Code:



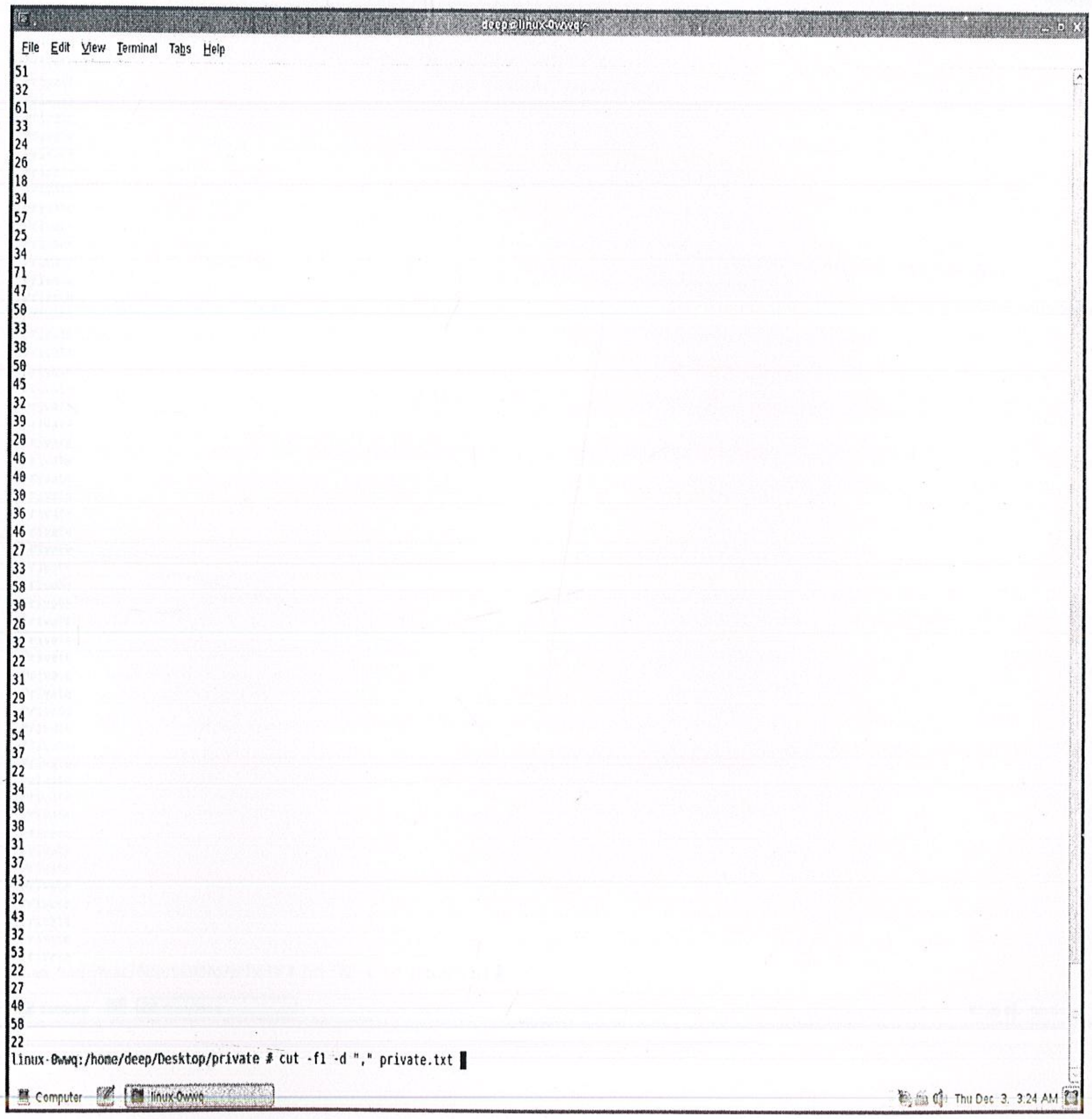
```
27, Private, 257302, Assoc-acdm, 12, Married-civ-spouse, Tech-support, Wife, White, Female, 0, 0, 38, United-States, <=50K
40, Private, 154374, HS-grad, 9, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 40, United-States, >50K
58, Private, 151910, HS-grad, 9, Widowed, Adm-clerical, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
22, Private, 201490, HS-grad, 9, Never-married, Adm-clerical, Own-child, White, Male, 0, 0, 20, United-States, <=50K
52, Self-emp-inc, 287927, HS-grad, 9, Married-civ-spouse, Exec-managerial, Wife, White, Female, 15024, 0, 40, United-States, >50K


linux-0wwq:/home/deep/Desktop # ls
adult.data.txt          input.php                                                                  new file 1       [redacted]
adultdata.txt           loaddata.png                                                               new file 1~      selfempinc.png
data1.html              [redacted]                                                                 package reading  [redacted]
data(2).php             localgov.png                                                               [redacted]       showtables.png
data.php                maindata.png                                                               privatetablepng  [redacted]
data.php~               MySQL  MySQL 3_23, 4_0, 4_1 Reference Manual  3_3_3 Loading Data into a Table.mht  pro.html  SuSE.desktop
[redacted]              [redacted]                                                                 pro.html~        [redacted]
GnomeOnlineHelp.desktop  neverworked.png                                                           relevent code    unknown.png
input (copy).php        new file                                                                   Screenshot1.png  [redacted]
linux-0wwq:/home/deep/Desktop # cd private
linux-0wwq:/home/deep/Desktop/private # cat table1.txt
#!/bin/bash
cp adult.data.txt adult.txt;
grep Private adult.txt > private.txt;
cut -f1 -d "," private.txt > private.1.txt;
cut -f2 -d "," private.txt > private.2.txt;
cut -f3 -d "," private.txt > private.3.txt;
cut -f4 -d "," private.txt > private.4.txt;
cut -f5 -d "," private.txt > private.5.txt;
cut -f6 -d "," private.txt > private.6.txt;
cut -f7 -d "," private.txt > private.7.txt;
cut -f8 -d "," private.txt > private.8.txt;
cut -f9 -d "," private.txt > private.9.txt;
cut -f10 -d "," private.txt > private.10.txt;
cut -f11 -d "," private.txt > private.11.txt;
cut -f12 -d "," private.txt > private.12.txt;
cut -f13 -d "," private.txt > private.13.txt;
cut -f14 -d "," private.txt > private.14.txt;
cut -f15 -d "," private.txt > private.15.txt;
paste private.1.txt private.2.txt > private.com1.txt;
paste private.com1.txt private.3.txt > private.com2.txt;
paste private.com2.txt private.4.txt > private.com3.txt;
paste private.com3.txt private.5.txt > private.com4.txt;
paste private.com4.txt private.6.txt > private.com5.txt;
paste private.com5.txt private.7.txt > private.com6.txt;
paste private.com6.txt private.8.txt > private.com7.txt;
paste private.com7.txt private.9.txt > private.com8.txt;
paste private.com8.txt private.10.txt > private.com9.txt;
paste private.com9.txt private.11.txt > private.com10.txt;
paste private.com10.txt private.12.txt > private.com11.txt;
paste private.com11.txt private.13.txt > private.com12.txt;
paste private.com12.txt private.14.txt > private.com13.txt;
paste private.com13.txt private.15.txt > private.com14.txt;


linux-0wwq:/home/deep/Desktop/private #
```

Private Data ‘grep’ped:

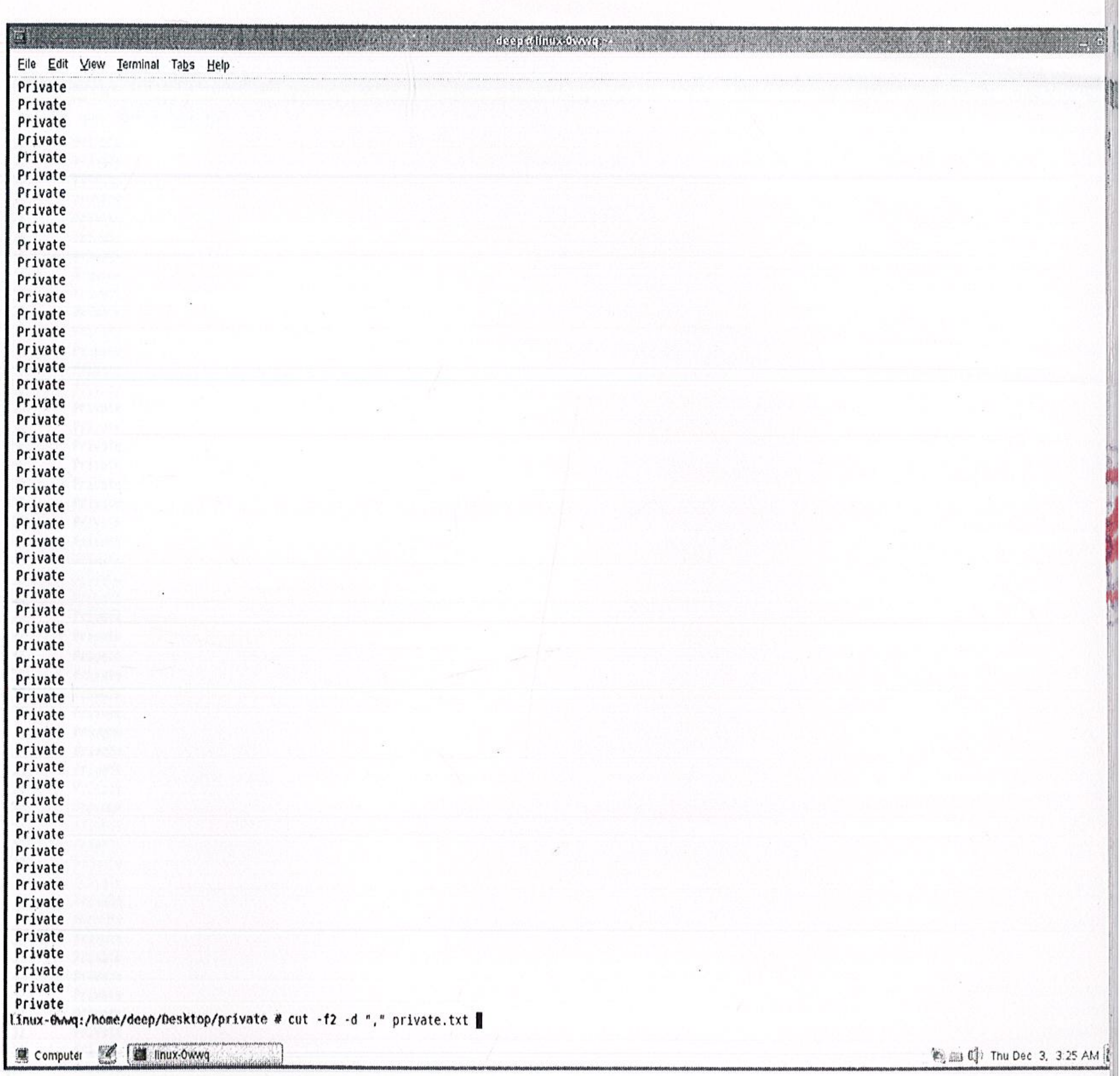


```
51, Private, 177669, Some-college, 10, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 60, United-States, <=50K
32, Private, 164190, Some-college, 10, Never-married, Exec-managerial, Own-child, White, Male, 0, 0, 40, United-States, <=50K
61, Private, 355645, HS-grad, 9, Married-civ-spouse, Sales, Husband, Black, Male, 0, 0, 40, United-States, <=50K
33, Private, 63079, HS-grad, 9, Divorced, Adm-clerical, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
24, Private, 381895, 11th, 7, Divorced, Machine-op-inspct, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
26, Private, 179010, Some-college, 10, Never-married, Craft-repair, Not-in-family, White, Male, 0, 0, 65, United-States, <=50K
18, Private, 436163, 11th, 7, Never-married, Prof-specialty, Own-child, White, Male, 0, 0, 20, United-States, <=50K
34, Private, 321709, HS-grad, 9, Never-married, Other-service, Not-in-family, White, Female, 0, 0, 28, United-States, <=50K
57, Private, 153918, HS-grad, 9, Married-civ-spouse, Transport-moving, Husband, White, Male, 0, 0, 40, United-States, <=50K
25, Private, 403788, HS-grad, 9, Never-married, Craft-repair, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
34, Private, 60567, 11th, 7, Divorced, Transport-moving, Unmarried, White, Male, 0, 880, 60, United-States, <=50K
71, Private, 138145, 9th, 5, Married-civ-spouse, Other-service, Husband, White, Male, 0, 0, 40, United-States, <=50K
47, Private, 312088, HS-grad, 9, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 40, United-States, <=50K
50, Private, 208630, Masters, 14, Divorced, Sales, Not-in-family, White, Female, 0, 0, 50, United-States, >50K
33, Private, 182401, 10th, 6, Never-married, Adm-clerical, Not-in-family, Black, Male, 0, 0, 40, United-States, <=50K
38, Private, 32916, Assoc-voc, 11, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 55, Unitea-States, >50K
50, Private, 302372, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 40, United-States, <=50K
45, Private, 155093, 10th, 6, Divorced, Other-service, Not-in-family, Black, Female, 0, 0, 38, Dominican-Republic, <=50K
32, Private, 192965, HS-grad, 9, Separated, Sales, Not-in-family, White, Female, 0, 0, 45, United-States, <=50K
39, Private, 107302, HS-grad, 9, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 45, ?, >50K
20, Private, 270436, HS-grad, 9, Never-married, Machine-op-inspct, Own-child, White, Male, 0, 0, 40, United-States, <=50K
46, Private, 42972, Masters, 14, Married-civ-spouse, Prof-specialty, Wife, White, Female, 0, 0, 22, United-States, >50K
40, Private, 142657, Assoc-voc, 11, Married-civ-spouse, Craft-repair, Husband, Black, Male, 0, 0, 45, United-States, <=50K
30, Private, 176175, Assoc-voc, 11, Divorced, Adm-clerical, Unmarried, White, Female, 0, 0, 24, United-States, <=50K
36, Private, 131459, 7th-8th, 4, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 40, United-States, <=50K
46, Private, 364548, Some-college, 10, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 48, United-States, >50K
27, Private, 177398, HS-grad, 9, Never-married, Other-service, Unmarried, White, Female, 0, 0, 64, United-States, <=50K
33, Private, 273243, HS-grad, 9, Married-civ-spouse, Craft-repair, Husband, Black, Male, 0, 0, 40, United-States, <=50K
58, Private, 147707, 11th, 7, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 40, United-States, <=50K
30, Private, 77266, HS-grad, 9, Divorced, Transport-moving, Not-in-family, White, Male, 0, 0, 55, United-States, <=50K
26, Private, 191648, Assoc-acdm, 12, Never-married, Machine-op-inspct, Other-relative, White, Female, 0, 0, 15, United-States, <=50K
32, Private, 211349, 10th, 6, Married-civ-spouse, Transport-moving, Husband, White, Male, 0, 0, 40, United-States, <=50K
22, Private, 203715, Some-college, 10, Never-married, Adm-clerical, Own-child, White, Male, 0, 0, 40, United-States, <=50K
31, Private, 292592, HS-grad, 9, Married-civ-spouse, Machine-op-inspct, Wife, White, Female, 0, 0, 40, United-States, <=50K
29, Private, 125976, HS-grad, 9, Separated, Sales, Unmarried, White, Female, 0, 0, 35, United-States, <=50K
34, Private, 204461, Doctorate, 16, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 60, United-States, >50K
54, Private, 337992, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, Asian-Pac-Islander, Male, 0, 0, 50, Japan, >50K
37, Private, 179137, Some-college, 10, Divorced, Adm-clerical, Unmarried, White, Female, 0, 0, 39, United-States, <=50K
22, Private, 325033, 12th, 8, Never-married, Protective-serv, Own-child, Black, Male, 0, 0, 35, United-States, <=50K
34, Private, 160216, Bachelors, 13, Never-married, Exec-managerial, Not-in-family, White, Female, 0, 0, 55, United-States, >50K
30, Private, 345898, HS-grad, 9, Never-married, Craft-repair, Not-in-family, Black, Male, 0, 0, 46, United-States, <=50K
38, Private, 139180, Bachelors, 13, Divorced, Prof-specialty, Unmarried, Black, Female, 15020, 0, 45, United-States, >50K
31, Private, 199655, Masters, 14, Divorced, Other-service, Not-in-family, Other, Female, 0, 0, 30, United-States, <=50K
37, Private, 198216, Assoc-acdm, 12, Divorced, Tech-support, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
43, Private, 260761, HS-grad, 9, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 40, Mexico, <=50K
32, Private, 34066, 10th, 6, Married-civ-spouse, Handlers-cleaners, Husband, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
43, Private, 84661, Assoc-voc, 11, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 45, United-States, <=50K
32, Private, 116138, Masters, 14, Never-married, Tech-support, Not-in-family, Asian-Pac-Islander, Male, 0, 0, 11, Taiwan, <=50K
53, Private, 321865, Masters, 14, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 40, United-States, >50K
22, Private, 310152, Some-college, 10, Never-married, Protective-serv, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
27, Private, 257302, Assoc-acdm, 12, Married-civ-spouse, Tech-support, Wife, White, Female, 0, 0, 38, United-States, <=50K
40, Private, 154374, HS-grad, 9, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 40, United-States, >50K
58, Private, 151910, HS-grad, 9, Widowed, Adm-clerical, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
22, Private, 201490, HS-grad, 9, Never-married, Adm-clerical, Own-child, White, Male, 0, 0, 20, United-States, <=50K
linux-0wwq:/home/deep/Desktop/private # grep Private adult.txt
```

Data 'Cut':

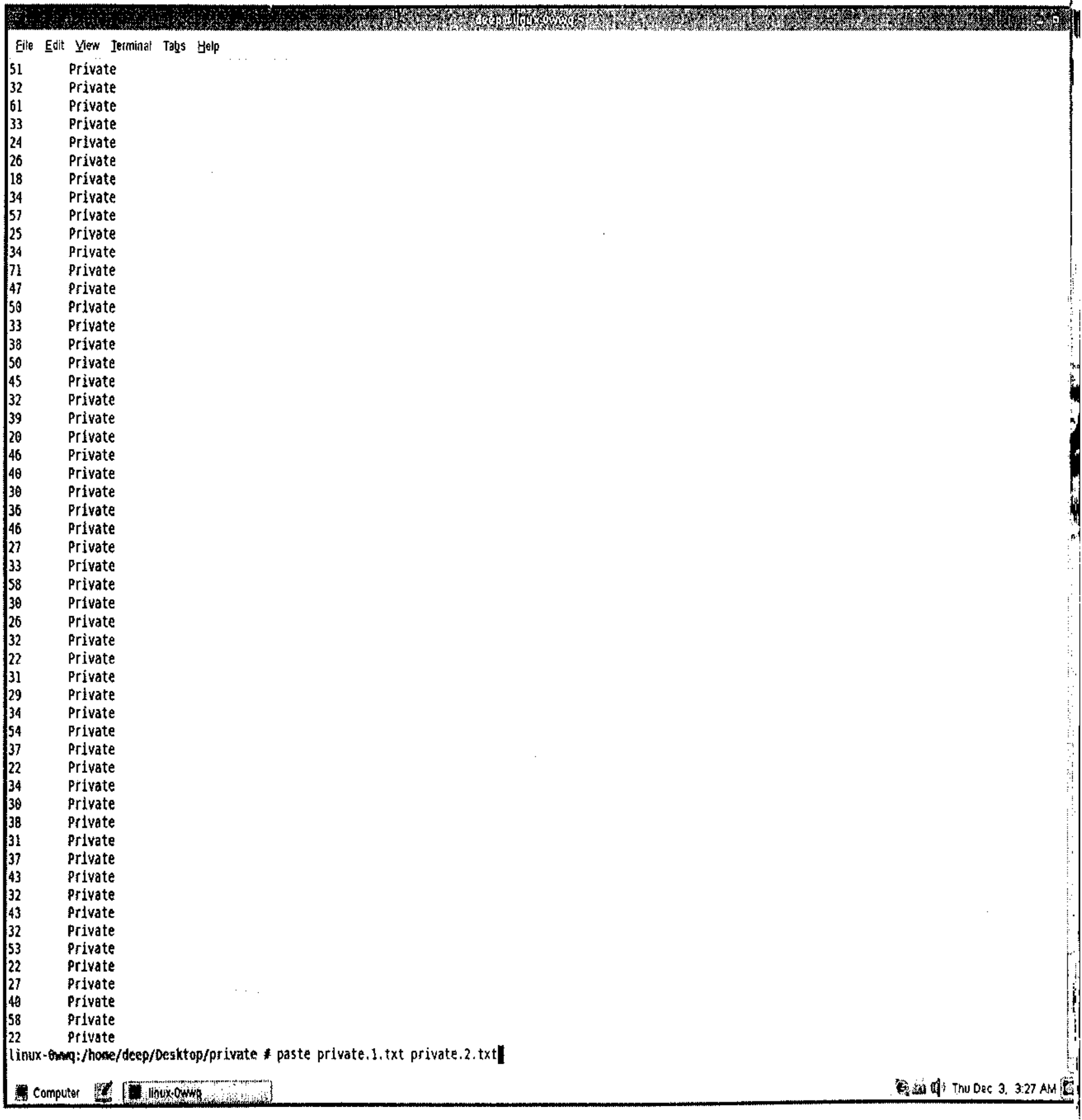


```
51
32
61
33
24
26
18
34
57
25
34
71
47
50
33
38
50
45
32
39
20
46
40
30
36
46
27
33
58
30
26
32
22
31
29
34
54
37
22
34
30
38
31
37
43
32
43
32
53
22
27
40
58
22
linux-0wwq:/home/deep/Desktop/private # cut -f1 -d "," private.txt
```



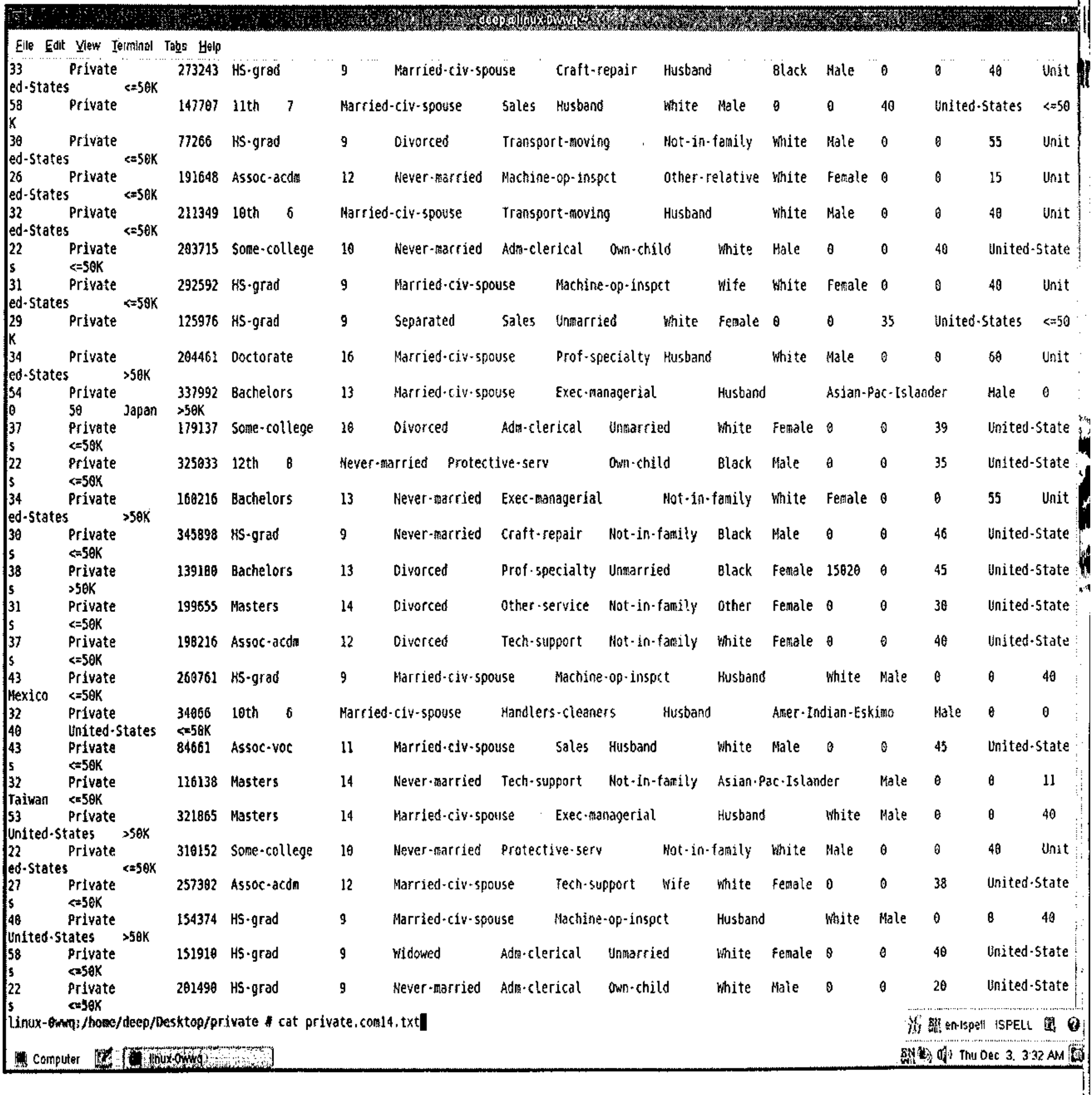
```
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
Private
linux-0wvq:/home/deep/Desktop/private # cut -f2 -d "," private.txt
```

Data 'Pasted'



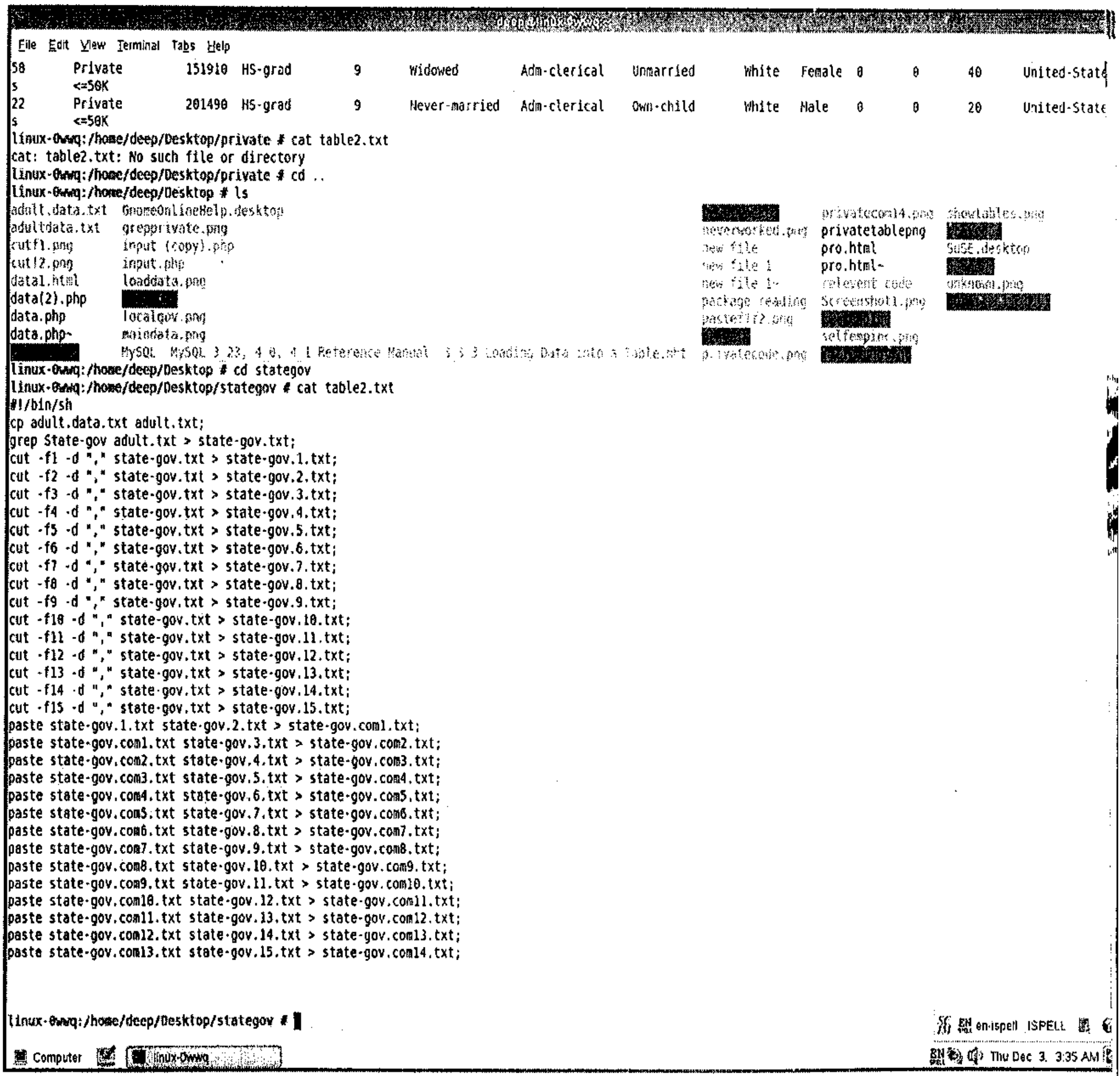
```
51      Private
32      Private
61      Private
33      Private
24      Private
26      Private
18      Private
34      Private
57      Private
25      Private
34      Private
71      Private
47      Private
50      Private
33      Private
38      Private
50      Private
45      Private
32      Private
39      Private
20      Private
46      Private
40      Private
30      Private
36      Private
46      Private
27      Private
33      Private
58      Private
30      Private
26      Private
32      Private
22      Private
31      Private
29      Private
34      Private
54      Private
37      Private
22      Private
34      Private
30      Private
38      Private
31      Private
37      Private
43      Private
32      Private
43      Private
32      Private
53      Private
22      Private
27      Private
40      Private
58      Private
22      Private
linux-0wwq:/home/deep/Desktop/private # paste private.1.txt private.2.txt
```

Private Data Concatenated:



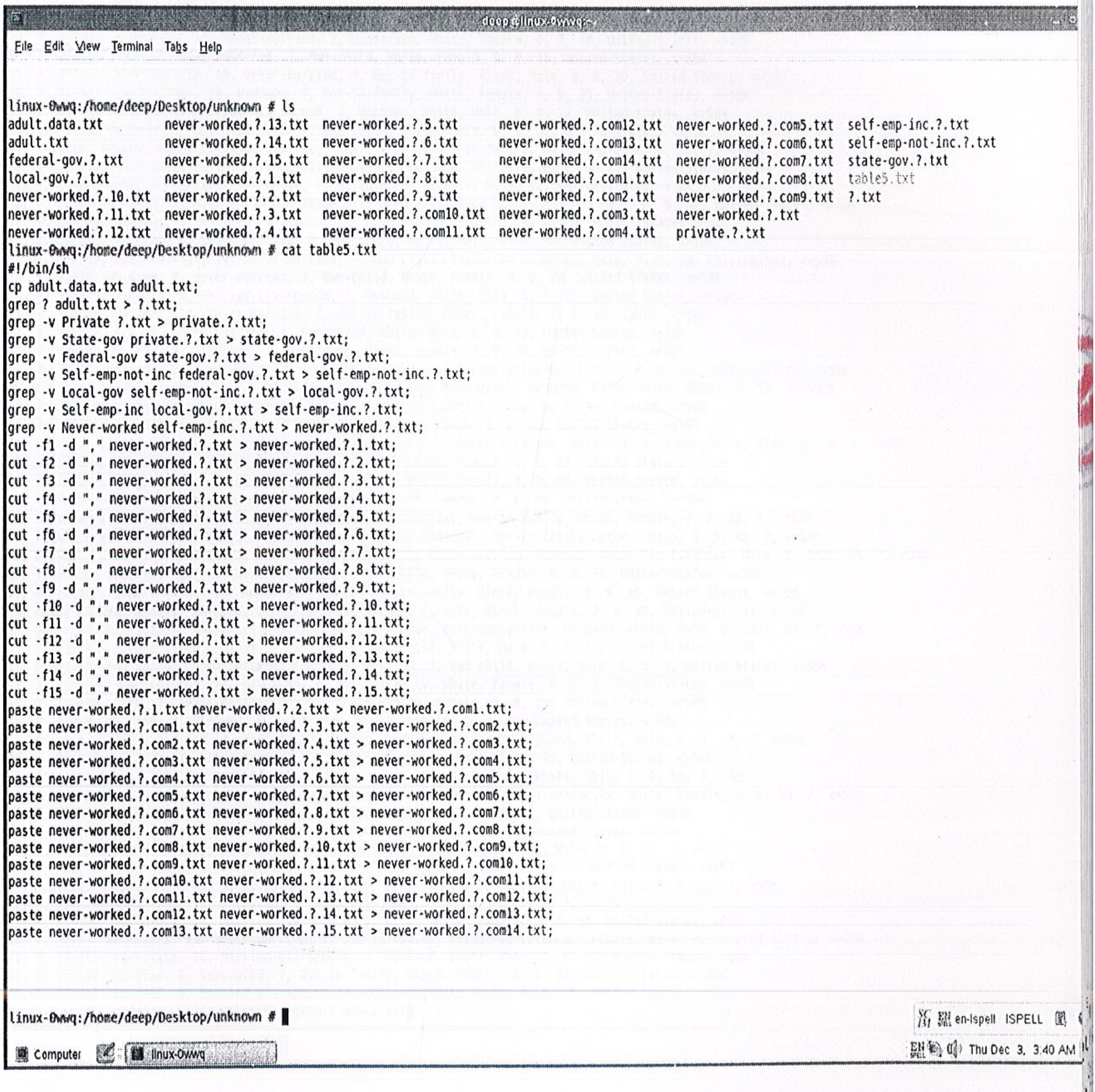
```
33      Private     273243  HS-grad       9       Married-civ-spouse     Craft-repair    Husband         Black   Male    0       0       40      Unit
ed-States       <=50K
58      Private     147707  11th    7       Married-civ-spouse      Sales   Husband         White   Male    0       0       40      United-States   <=50
K
30      Private     77266   HS-grad       9       Divorced        Transport-moving        Not-in-family   White   Male    0       0       55      Unit
ed-States       <=50K
26      Private     191648  Assoc-acdm    12      Never-married   Machine-op-inspct       Other-relative  White   Female  0       0       15      Unit
ed-States       <=50K
32      Private     211349  10th    6       Married-civ-spouse      Transport-moving        Husband         White   Male    0       0       40      Unit
ed-States       <=50K
22      Private     203715  Some-college  10      Never-married   Adm-clerical    Own-child       White   Male    0       0       40      United-State
s       <=50K
31      Private     292592  HS-grad       9       Married-civ-spouse      Machine-op-inspct       Wife    White   Female  0       0       40      Unit
ed-States       <=50K
29      Private     125976  HS-grad       9       Separated       Sales   Unmarried       White   Female  0       0       35      United-States   <=50
K
34      Private     204461  Doctorate     16      Married-civ-spouse      Prof-specialty  Husband         White   Male    0       0       60      Unit
ed-States       >50K
54      Private     337992  Bachelors     13      Married-civ-spouse      Exec-managerial         Husband         Asian-Pac-Islander      Male    0
0       50      Japan   >50K
37      Private     179137  Some-college  10      Divorced        Adm-clerical    Unmarried       White   Female  0       0       39      United-State
s       <=50K
22      Private     325033  12th    8       Never-married   Protective-serv         Own-child       Black   Male    0       0       35      United-State
s       <=50K
34      Private     160216  Bachelors     13      Never-married   Exec-managerial         Not-in-family   White   Female  0       0       55      Unit
ed-States       >50K
30      Private     345898  HS-grad       9       Never-married   Craft-repair    Not-in-family   Black   Male    0       0       46      United-State
s       <=50K
38      Private     139180  Bachelors     13      Divorced        Prof-specialty  Unmarried       Black   Female  15020   0       45      United-State
s       >50K
31      Private     199655  Masters       14      Divorced        Other-service   Not-in-family   Other   Female  0       0       30      United-State
s       <=50K
37      Private     198216  Assoc-acdm    12      Divorced        Tech-support    Not-in-family   White   Female  0       0       40      United-State
s       <=50K
43      Private     260761  HS-grad       9       Married-civ-spouse      Machine-op-inspct       Husband         White   Male    0       0       40
Mexico  <=50K
32      Private     34066   10th    6       Married-civ-spouse      Handlers-cleaners       Husband         Amer-Indian-Eskimo      Male    0       0
40      United-States   <=50K
43      Private     84661   Assoc-voc     11      Married-civ-spouse      Sales   Husband         White   Male    0       0       45      United-State
s       <=50K
32      Private     116138  Masters       14      Never-married   Tech-support    Not-in-family   Asian-Pac-Islander      Male    0       0       11
Taiwan  <=50K
53      Private     321865  Masters       14      Married-civ-spouse      Exec-managerial         Husband         White   Male    0       0       40
United-States   >50K
22      Private     310152  Some-college  10      Never-married   Protective-serv         Not-in-family   White   Male    0       0       40      Unit
ed-States       <=50K
27      Private     257302  Assoc-acdm    12      Married-civ-spouse      Tech-support    Wife    White   Female  0       0       38      United-State
s       <=50K
40      Private     154374  HS-grad       9       Married-civ-spouse      Machine-op-inspct       Husband         White   Male    0       0       40
United-States   >50K
58      Private     151910  HS-grad       9       Widowed         Adm-clerical    Unmarried       White   Female  0       0       40      United-State
s       <=50K
22      Private     201490  HS-grad       9       Never-married   Adm-clerical    Own-child       White   Male    0       0       20      United-State
s       <=50K
linux-0wwq:/home/deep/Desktop/private # cat private.com14.txt
```

State gov. data:



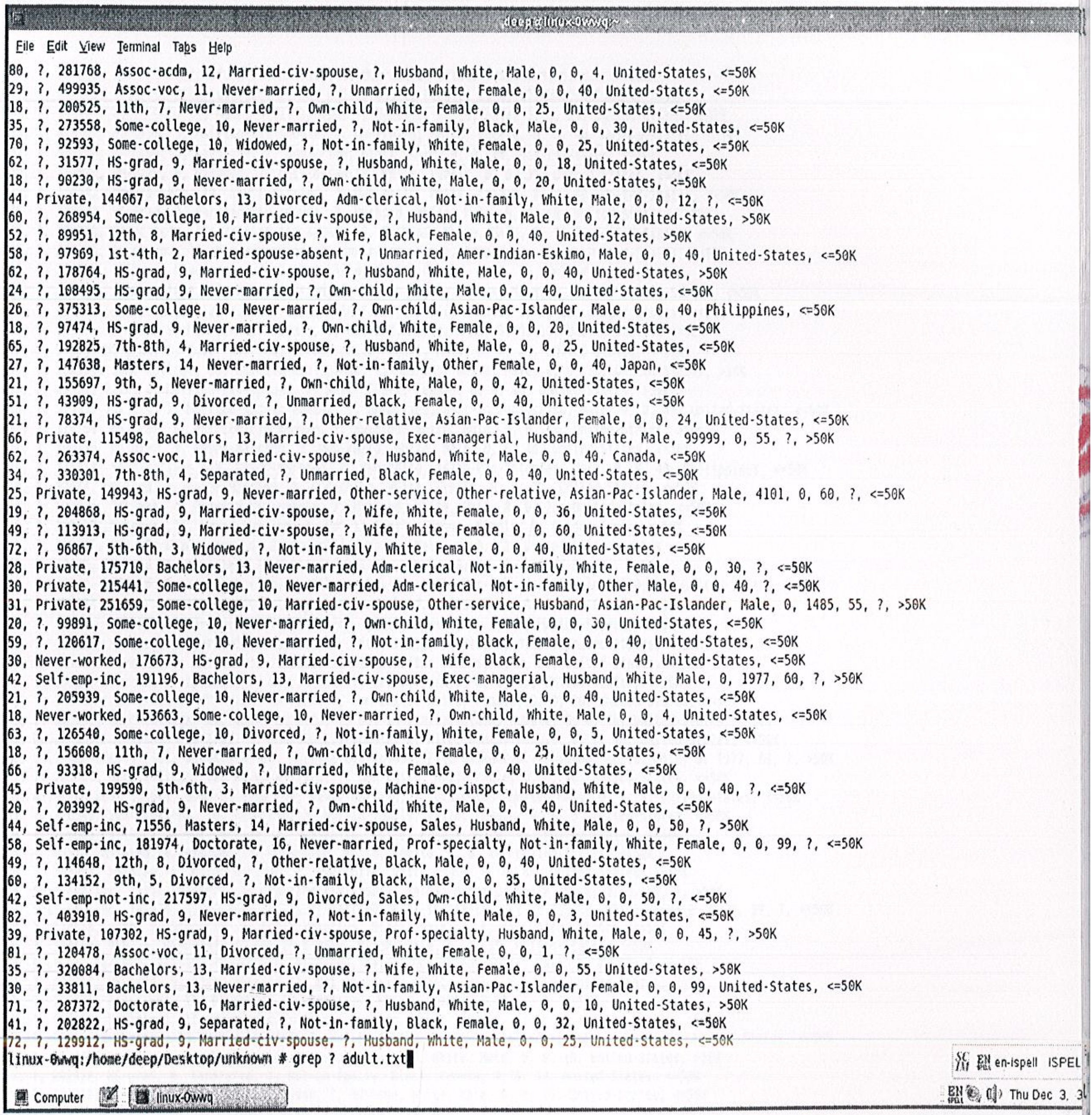
```
58      Private         151910  HS-grad         9       Widowed         Adm-clerical    Unmarried       White   Female  0       0       40      United-State
s       <=50K
22      Private         201490  HS-grad         9       Never-married   Adm-clerical    Own-child       White   Male    0       0       20      United-State
s       <=50K
linux-0wwq:/home/deep/Desktop/private # cat table2.txt
cat: table2.txt: No such file or directory
linux-0wwq:/home/deep/Desktop/private # cd ..
linux-0wwq:/home/deep/Desktop # ls
adult.data.txt  GnomeOnlineHelp.desktop                                                       [redacted]       privatecom14.png  showtables.png
adultdata.txt   grepprivate.png                                                               neverworked.png  privatetablepng   [redacted]
cutf1.png       input (copy).php                                                              new file         pro.html          SuSE.desktop
cutf2.png       input.php                                                                     new file 1       pro.html~         [redacted]
data1.html      loaddata.png                                                                  new file 1~      relevent code     unknown.png
data(2).php     [redacted]                                                                    package reading  Screenshot1.png   [redacted]
data.php        localgov.png                                                                  pastef1f2.png    [redacted]
data.php~       maindata.png                                                                  [redacted]       selfempinc.png
[redacted]      MySQL  MySQL 3.23, 4.0, 4.1 Reference Manual  3.3.3 Loading Data into a Table.mht  privatecode.png  [redacted]
linux-0wwq:/home/deep/Desktop # cd stategov
linux-0wwq:/home/deep/Desktop/stategov # cat table2.txt
#!/bin/sh
cp adult.data.txt adult.txt;
grep State-gov adult.txt > state-gov.txt;
cut -f1 -d "," state-gov.txt > state-gov.1.txt;
cut -f2 -d "," state-gov.txt > state-gov.2.txt;
cut -f3 -d "," state-gov.txt > state-gov.3.txt;
cut -f4 -d "," state-gov.txt > state-gov.4.txt;
cut -f5 -d "," state-gov.txt > state-gov.5.txt;
cut -f6 -d "," state-gov.txt > state-gov.6.txt;
cut -f7 -d "," state-gov.txt > state-gov.7.txt;
cut -f8 -d "," state-gov.txt > state-gov.8.txt;
cut -f9 -d "," state-gov.txt > state-gov.9.txt;
cut -f10 -d "," state-gov.txt > state-gov.10.txt;
cut -f11 -d "," state-gov.txt > state-gov.11.txt;
cut -f12 -d "," state-gov.txt > state-gov.12.txt;
cut -f13 -d "," state-gov.txt > state-gov.13.txt;
cut -f14 -d "," state-gov.txt > state-gov.14.txt;
cut -f15 -d "," state-gov.txt > state-gov.15.txt;
paste state-gov.1.txt state-gov.2.txt > state-gov.com1.txt;
paste state-gov.com1.txt state-gov.3.txt > state-gov.com2.txt;
paste state-gov.com2.txt state-gov.4.txt > state-gov.com3.txt;
paste state-gov.com3.txt state-gov.5.txt > state-gov.com4.txt;
paste state-gov.com4.txt state-gov.6.txt > state-gov.com5.txt;
paste state-gov.com5.txt state-gov.7.txt > state-gov.com6.txt;
paste state-gov.com6.txt state-gov.8.txt > state-gov.com7.txt;
paste state-gov.com7.txt state-gov.9.txt > state-gov.com8.txt;
paste state-gov.com8.txt state-gov.10.txt > state-gov.com9.txt;
paste state-gov.com9.txt state-gov.11.txt > state-gov.com10.txt;
paste state-gov.com10.txt state-gov.12.txt > state-gov.com11.txt;
paste state-gov.com11.txt state-gov.13.txt > state-gov.com12.txt;
paste state-gov.com12.txt state-gov.14.txt > state-gov.com13.txt;
paste state-gov.com13.txt state-gov.15.txt > state-gov.com14.txt;


linux-0wwq:/home/deep/Desktop/stategov #
```

Unknown-labeled Data:



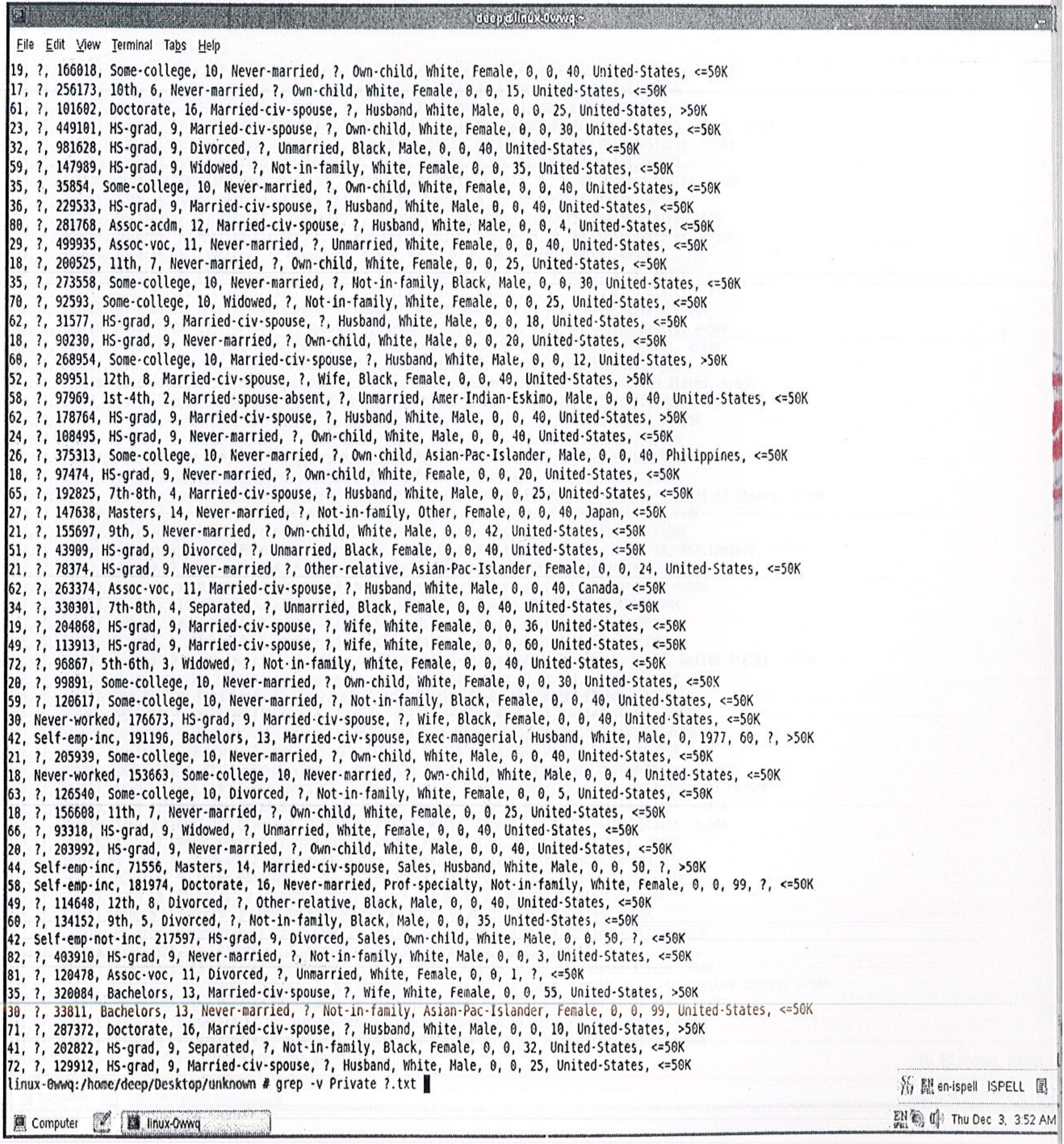
```
linux-0wwq:/home/deep/Desktop/unknown # ls
adult.data.txt         never-worked.?.13.txt  never-worked.?.5.txt       never-worked.?.com12.txt  never-worked.?.com5.txt  self-emp-inc.?.txt
adult.txt              never-worked.?.14.txt  never-worked.?.6.txt       never-worked.?.com13.txt  never-worked.?.com6.txt  self-emp-not-inc.?.txt
federal-gov.?.txt      never-worked.?.15.txt  never-worked.?.7.txt       never-worked.?.com14.txt  never-worked.?.com7.txt  state-gov.?.txt
local-gov.?.txt        never-worked.?.1.txt   never-worked.?.8.txt       never-worked.?.com1.txt   never-worked.?.com8.txt  table5.txt
never-worked.?.10.txt  never-worked.?.2.txt   never-worked.?.9.txt       never-worked.?.com2.txt   never-worked.?.com9.txt  ?.txt
never-worked.?.11.txt  never-worked.?.3.txt   never-worked.?.com10.txt   never-worked.?.com3.txt   never-worked.?.txt
never-worked.?.12.txt  never-worked.?.4.txt   never-worked.?.com11.txt   never-worked.?.com4.txt   private.?.txt
linux-0wwq:/home/deep/Desktop/unknown # cat table5.txt
#!/bin/sh
cp adult.data.txt adult.txt;
grep ? adult.txt > ?.txt;
grep -v Private ?.txt > private.?.txt;
grep -v State-gov private.?.txt > state-gov.?.txt;
grep -v Federal-gov state-gov.?.txt > federal-gov.?.txt;
grep -v Self-emp-not-inc federal-gov.?.txt > self-emp-not-inc.?.txt;
grep -v Local-gov self-emp-not-inc.?.txt > local-gov.?.txt;
grep -v Self-emp-inc local-gov.?.txt > self-emp-inc.?.txt;
grep -v Never-worked self-emp-inc.?.txt > never-worked.?.txt;
cut -f1 -d "," never-worked.?.txt > never-worked.?.1.txt;
cut -f2 -d "," never-worked.?.txt > never-worked.?.2.txt;
cut -f3 -d "," never-worked.?.txt > never-worked.?.3.txt;
cut -f4 -d "," never-worked.?.txt > never-worked.?.4.txt;
cut -f5 -d "," never-worked.?.txt > never-worked.?.5.txt;
cut -f6 -d "," never-worked.?.txt > never-worked.?.6.txt;
cut -f7 -d "," never-worked.?.txt > never-worked.?.7.txt;
cut -f8 -d "," never-worked.?.txt > never-worked.?.8.txt;
cut -f9 -d "," never-worked.?.txt > never-worked.?.9.txt;
cut -f10 -d "," never-worked.?.txt > never-worked.?.10.txt;
cut -f11 -d "," never-worked.?.txt > never-worked.?.11.txt;
cut -f12 -d "," never-worked.?.txt > never-worked.?.12.txt;
cut -f13 -d "," never-worked.?.txt > never-worked.?.13.txt;
cut -f14 -d "," never-worked.?.txt > never-worked.?.14.txt;
cut -f15 -d "," never-worked.?.txt > never-worked.?.15.txt;
paste never-worked.?.1.txt never-worked.?.2.txt > never-worked.?.com1.txt;
paste never-worked.?.com1.txt never-worked.?.3.txt > never-worked.?.com2.txt;
paste never-worked.?.com2.txt never-worked.?.4.txt > never-worked.?.com3.txt;
paste never-worked.?.com3.txt never-worked.?.5.txt > never-worked.?.com4.txt;
paste never-worked.?.com4.txt never-worked.?.6.txt > never-worked.?.com5.txt;
paste never-worked.?.com5.txt never-worked.?.7.txt > never-worked.?.com6.txt;
paste never-worked.?.com6.txt never-worked.?.8.txt > never-worked.?.com7.txt;
paste never-worked.?.com7.txt never-worked.?.9.txt > never-worked.?.com8.txt;
paste never-worked.?.com8.txt never-worked.?.10.txt > never-worked.?.com9.txt;
paste never-worked.?.com9.txt never-worked.?.11.txt > never-worked.?.com10.txt;
paste never-worked.?.com10.txt never-worked.?.12.txt > never-worked.?.com11.txt;
paste never-worked.?.com11.txt never-worked.?.13.txt > never-worked.?.com12.txt;
paste never-worked.?.com12.txt never-worked.?.14.txt > never-worked.?.com13.txt;
paste never-worked.?.com13.txt never-worked.?.15.txt > never-worked.?.com14.txt;

linux-0wwq:/home/deep/Desktop/unknown # ▌
```

'grep'ing Adult data:



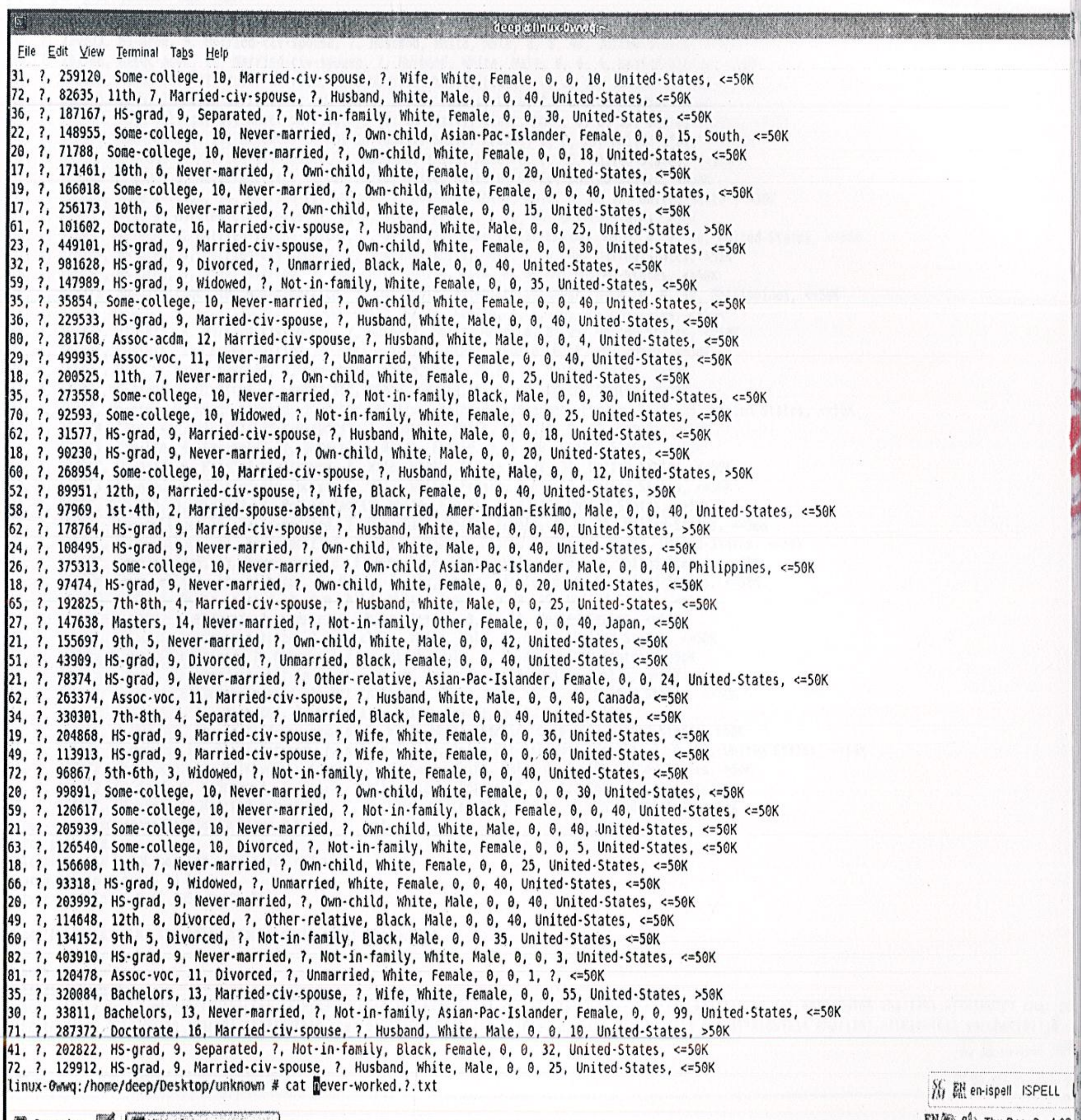
```
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
44, Private, 144067, Bachelors, 13, Divorced, Adm-clerical, Not-in-family, White, Male, 0, 0, 12, ?, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
66, Private, 115498, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 99999, 0, 55, ?, >50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
25, Private, 149943, HS-grad, 9, Never-married, Other-service, Other-relative, Asian-Pac-Islander, Male, 4101, 0, 60, ?, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
28, Private, 175710, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Female, 0, 0, 30, ?, <=50K
30, Private, 215441, Some-college, 10, Never-married, Adm-clerical, Not-in-family, Other, Male, 0, 0, 40, ?, <=50K
31, Private, 251659, Some-college, 10, Married-civ-spouse, Other-service, Husband, Asian-Pac-Islander, Male, 0, 1485, 55, ?, >50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
30, Never-worked, 176673, HS-grad, 9, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, <=50K
42, Self-emp-inc, 191196, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 1977, 60, ?, >50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
18, Never-worked, 153663, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 4, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
45, Private, 199590, 5th-6th, 3, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 40, ?, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
44, Self-emp-inc, 71556, Masters, 14, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 50, ?, >50K
58, Self-emp-inc, 181974, Doctorate, 16, Never-married, Prof-specialty, Not-in-family, White, Female, 0, 0, 99, ?, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
42, Self-emp-not-inc, 217597, HS-grad, 9, Divorced, Sales, Own-child, White, Male, 0, 0, 50, ?, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
39, Private, 107302, HS-grad, 9, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 45, ?, >50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # grep ? adult.txt
```

Eradicating ‘Private’:



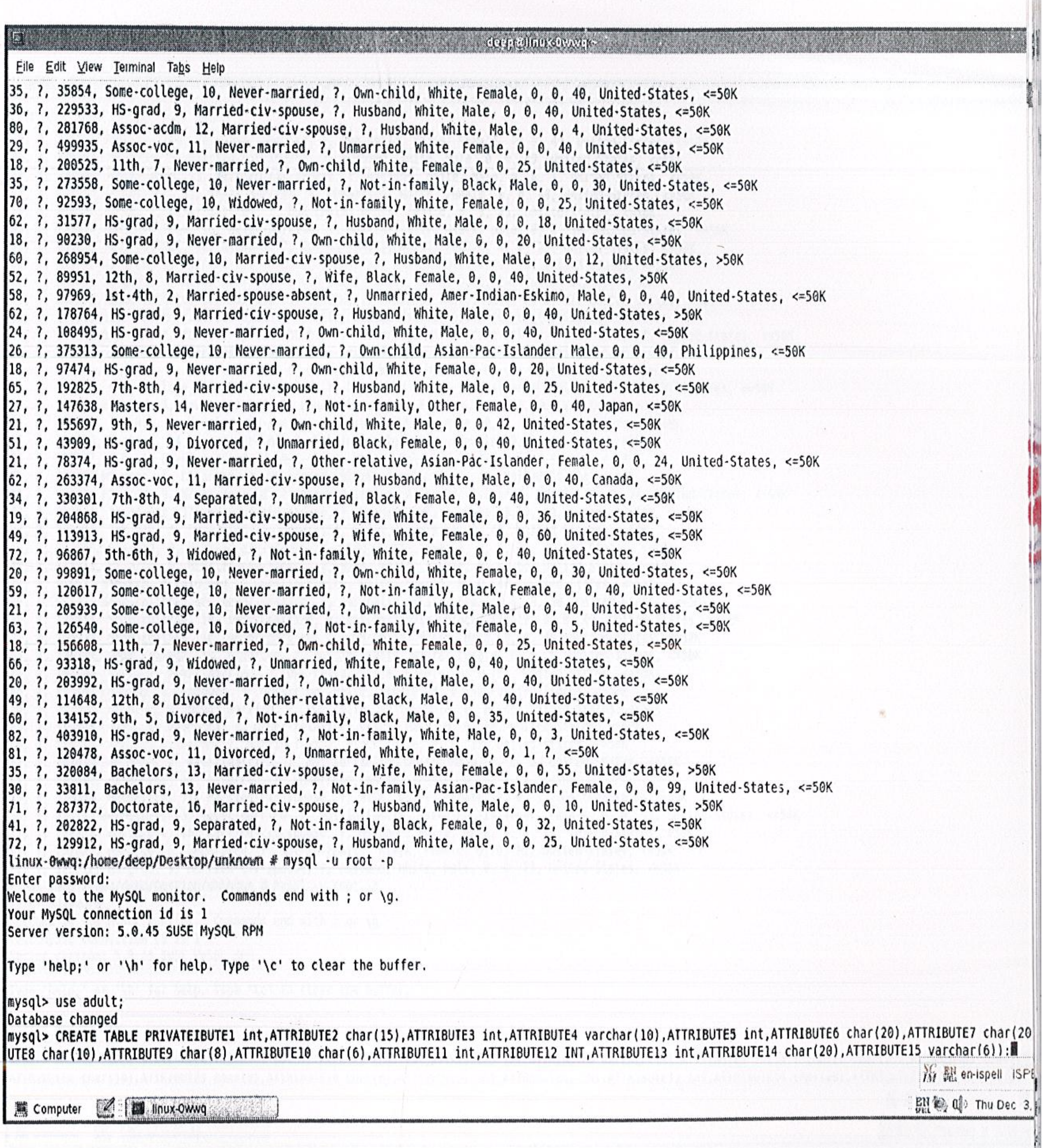
```
19, ?, 166018, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
17, ?, 256173, 10th, 6, Never-married, ?, Own-child, White, Female, 0, 0, 15, United-States, <=50K
61, ?, 101602, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, >50K
23, ?, 449101, HS-grad, 9, Married-civ-spouse, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
32, ?, 981628, HS-grad, 9, Divorced, ?, Unmarried, Black, Male, 0, 0, 40, United-States, <=50K
59, ?, 147989, HS-grad, 9, Widowed, ?, Not-in-family, White, Female, 0, 0, 35, United-States, <=50K
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
30, Never-worked, 176673, HS-grad, 9, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, <=50K
42, Self-emp-inc, 191196, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 1977, 60, ?, >50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
18, Never-worked, 153663, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 4, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
44, Self-emp-inc, 71556, Masters, 14, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 50, ?, >50K
58, Self-emp-inc, 181974, Doctorate, 16, Never-married, Prof-specialty, Not-in-family, White, Female, 0, 0, 99, ?, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
42, Self-emp-not-inc, 217597, HS-grad, 9, Divorced, Sales, Own-child, White, Male, 0, 0, 50, ?, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # grep -v Private ?.txt
```

After removal:



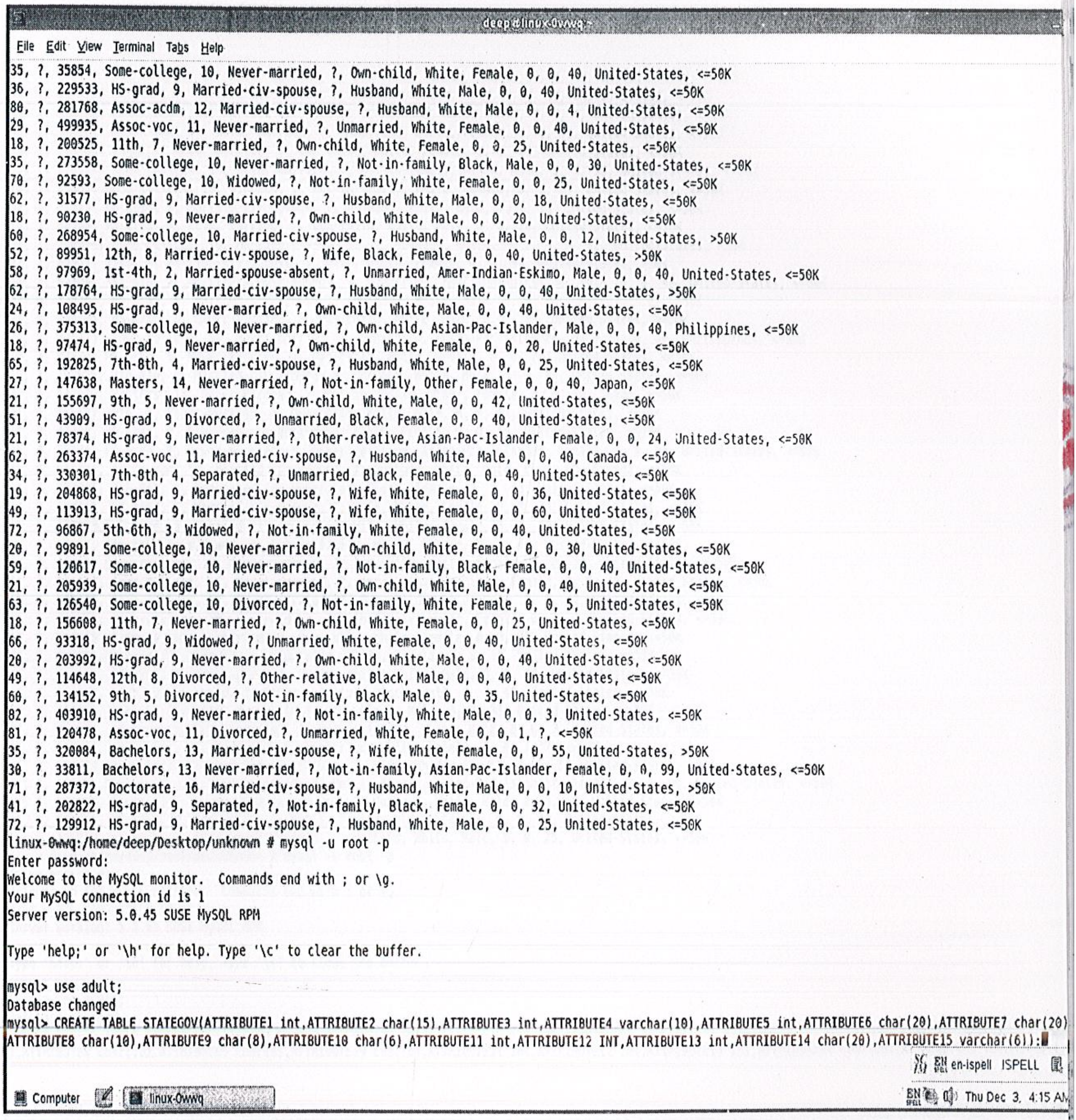
```
31, ?, 259120, Some-college, 10, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 10, United-States, <=50K
72, ?, 82635, 11th, 7, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
36, ?, 187167, HS-grad, 9, Separated, ?, Not-in-family, White, Female, 0, 0, 30, United-States, <=50K
22, ?, 148955, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Female, 0, 0, 15, South, <=50K
20, ?, 71788, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 18, United-States, <=50K
17, ?, 171461, 10th, 6, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
19, ?, 166018, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
17, ?, 256173, 10th, 6, Never-married, ?, Own-child, White, Female, 0, 0, 15, United-States, <=50K
61, ?, 101602, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, >50K
23, ?, 449101, HS-grad, 9, Married-civ-spouse, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
32, ?, 981628, HS-grad, 9, Divorced, ?, Unmarried, Black, Male, 0, 0, 40, United-States, <=50K
59, ?, 147989, HS-grad, 9, Widowed, ?, Not-in-family, White, Female, 0, 0, 35, United-States, <=50K
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # cat never-worked.?.txt
```

Private table:



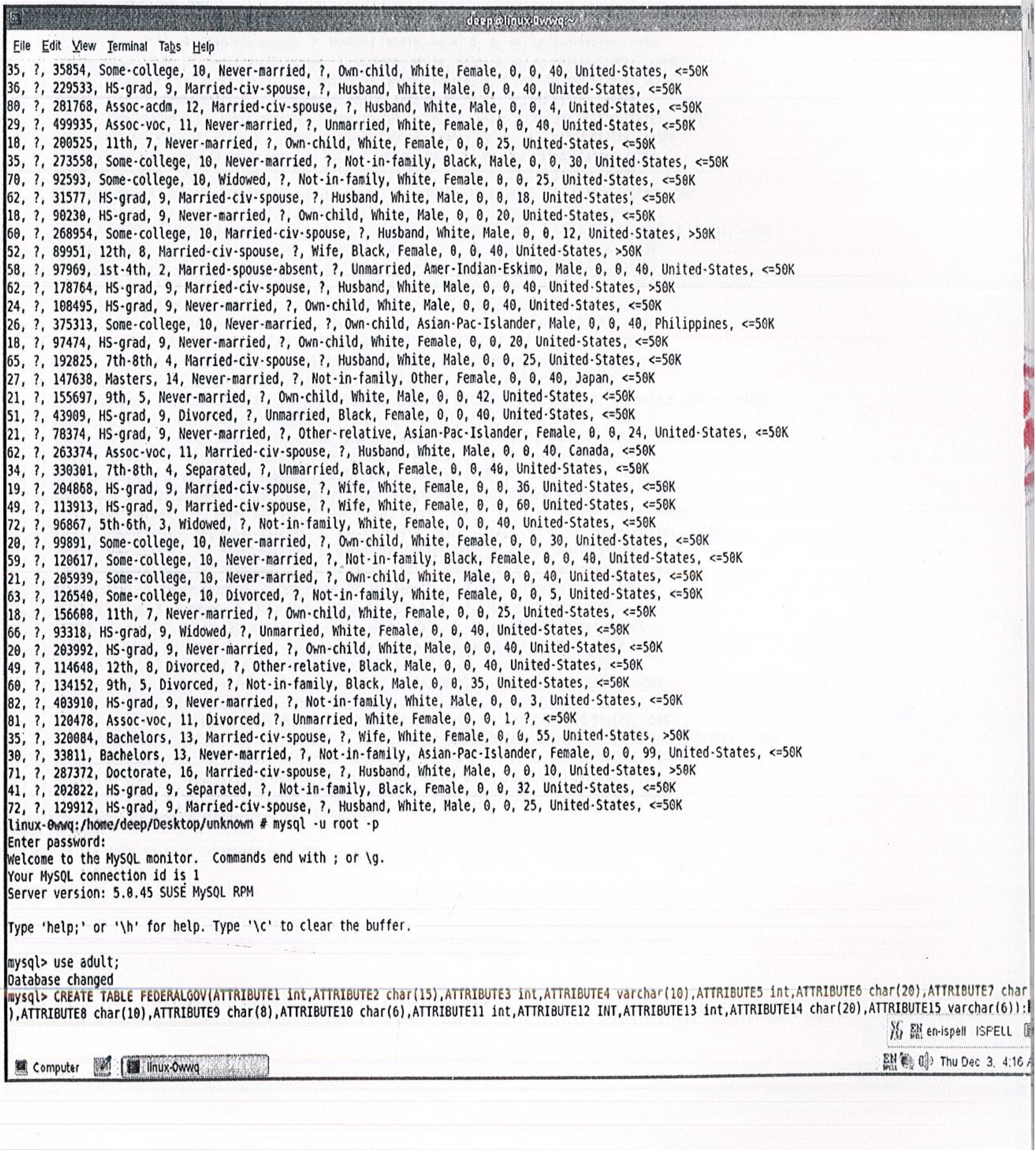
```
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 6, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, e, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.45 SUSE MySQL RPM

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> use adult;
Database changed
mysql> CREATE TABLE PRIVATEIBUTE1 int,ATTRIBUTE2 char(15),ATTRIBUTE3 int,ATTRIBUTE4 varchar(10),ATTRIBUTE5 int,ATTRIBUTE6 char(20),ATTRIBUTE7 char(20
UTE8 char(10),ATTRIBUTE9 char(8),ATTRIBUTE10 char(6),ATTRIBUTE11 int,ATTRIBUTE12 INT,ATTRIBUTE13 int,ATTRIBUTE14 char(20),ATTRIBUTE15 varchar(6)):
```

Stategov. Table:



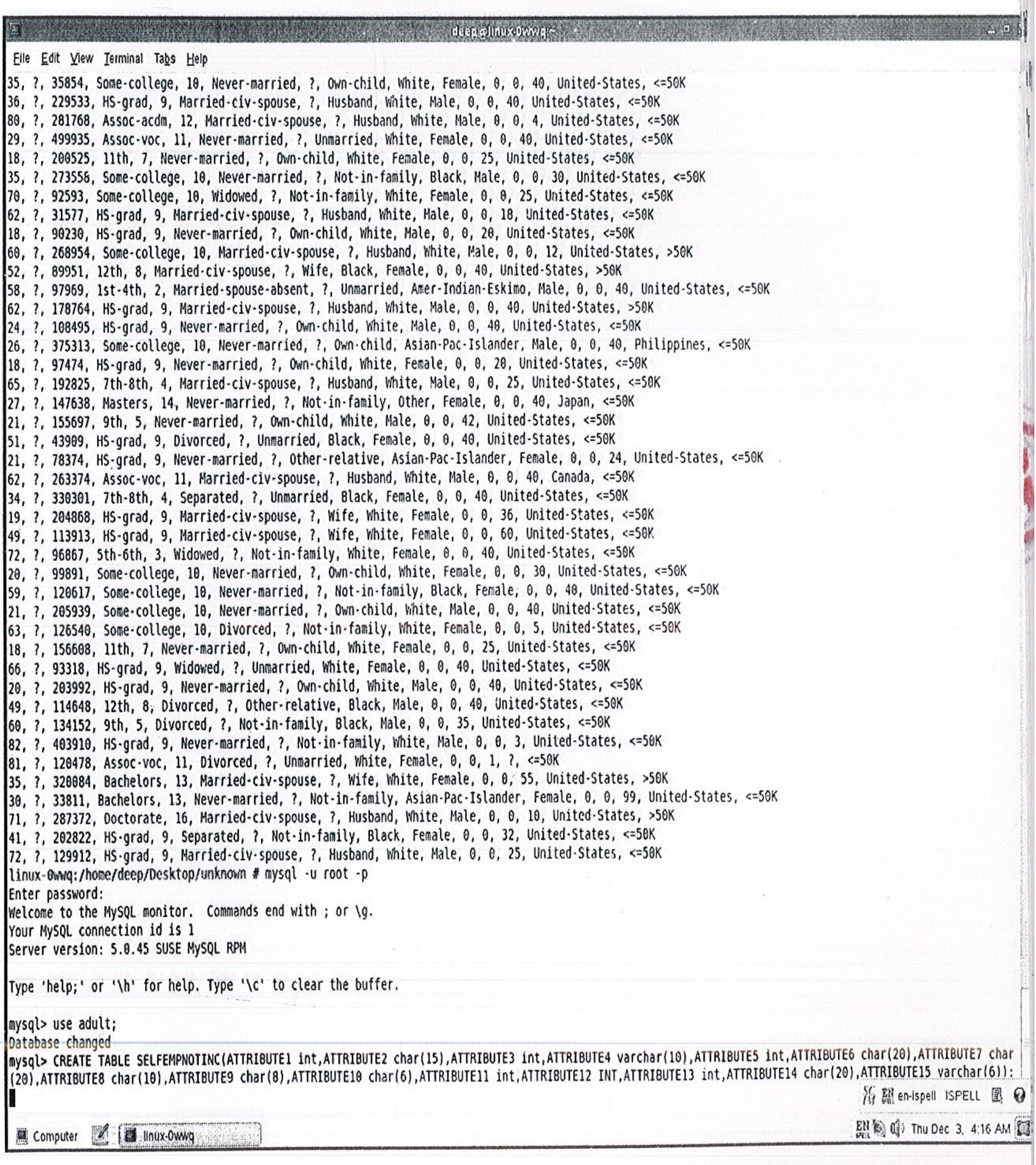
```
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.45 SUSE MySQL RPM

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> use adult;
Database changed
mysql> CREATE TABLE STATEGOV(ATTRIBUTE1 int,ATTRIBUTE2 char(15),ATTRIBUTE3 int,ATTRIBUTE4 varchar(10),ATTRIBUTE5 int,ATTRIBUTE6 char(20),ATTRIBUTE7 char(20)
ATTRIBUTE8 char(10),ATTRIBUTE9 char(8),ATTRIBUTE10 char(6),ATTRIBUTE11 int,ATTRIBUTE12 INT,ATTRIBUTE13 int,ATTRIBUTE14 char(20),ATTRIBUTE15 varchar(6));
```

Federal govt. data:

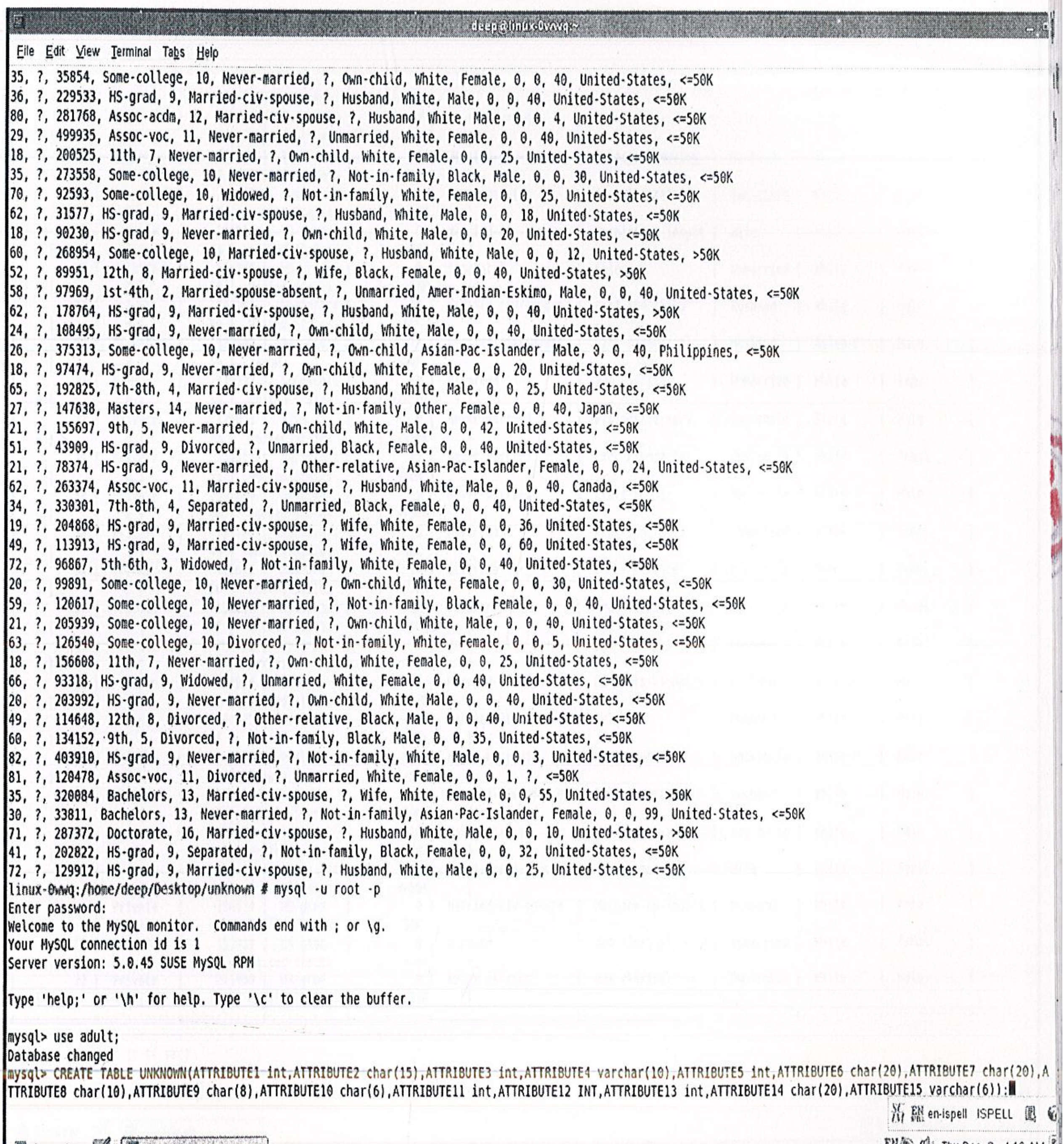


```
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.45 SUSE MySQL RPM

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> use adult;
Database changed
mysql> CREATE TABLE FEDERALGOV(ATTRIBUTE1 int,ATTRIBUTE2 char(15),ATTRIBUTE3 int,ATTRIBUTE4 varchar(10),ATTRIBUTE5 int,ATTRIBUTE6 char(20),ATTRIBUTE7 char
),ATTRIBUTE8 char(10),ATTRIBUTE9 char(8),ATTRIBUTE10 char(6),ATTRIBUTE11 int,ATTRIBUTE12 INT,ATTRIBUTE13 int,ATTRIBUTE14 char(20),ATTRIBUTE15 varchar(6)):
```

Database changed:

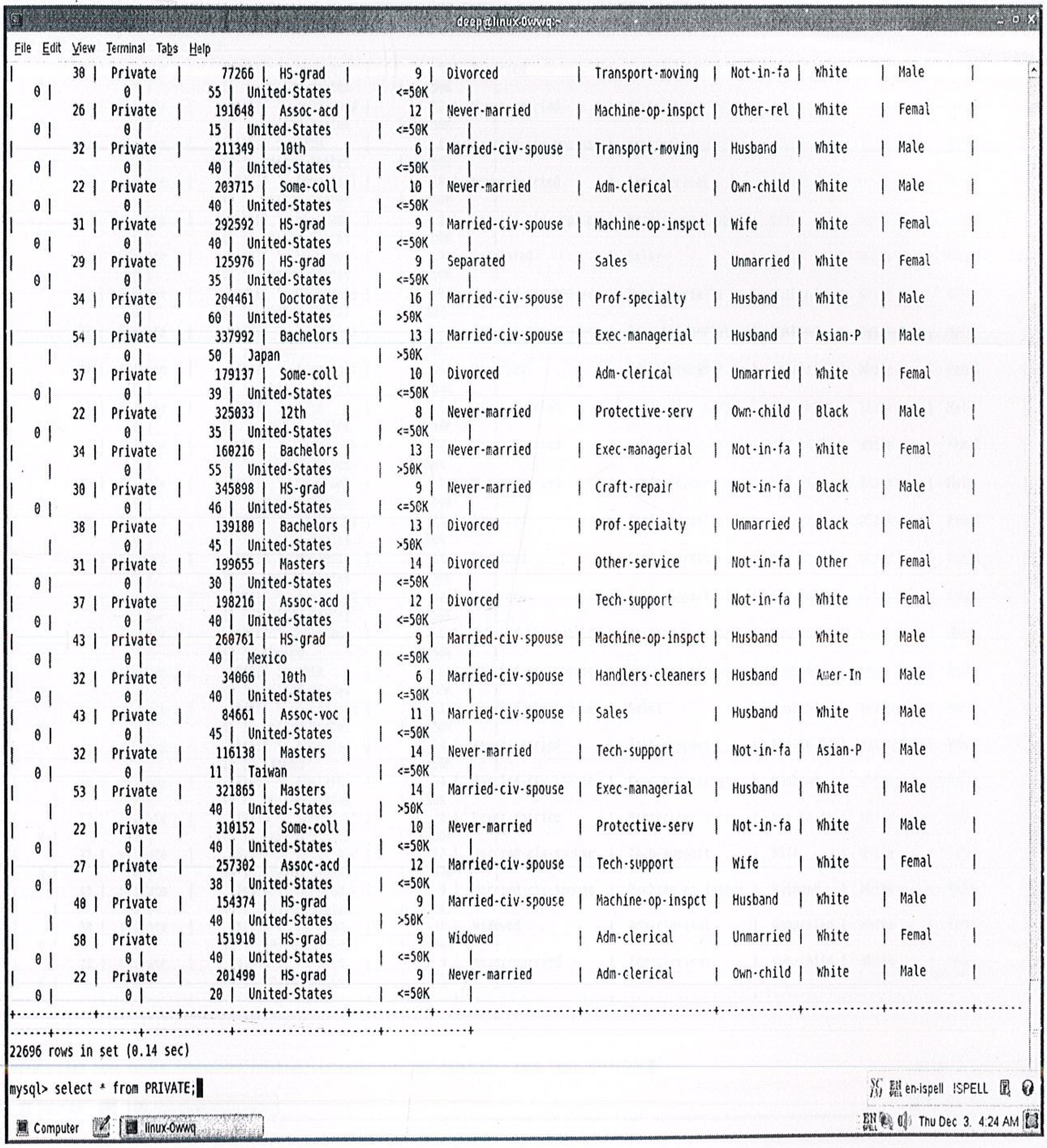


```
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273556, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.45 SUSE MySQL RPM

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> use adult;
Database changed
mysql> CREATE TABLE SELFEMPNOTINC(ATTRIBUTE1 int,ATTRIBUTE2 char(15),ATTRIBUTE3 int,ATTRIBUTE4 varchar(10),ATTRIBUTE5 int,ATTRIBUTE6 char(20),ATTRIBUTE7 char
(20),ATTRIBUTE8 char(10),ATTRIBUTE9 char(8),ATTRIBUTE10 char(6),ATTRIBUTE11 int,ATTRIBUTE12 INT,ATTRIBUTE13 int,ATTRIBUTE14 char(20),ATTRIBUTE15 varchar(6)):
```



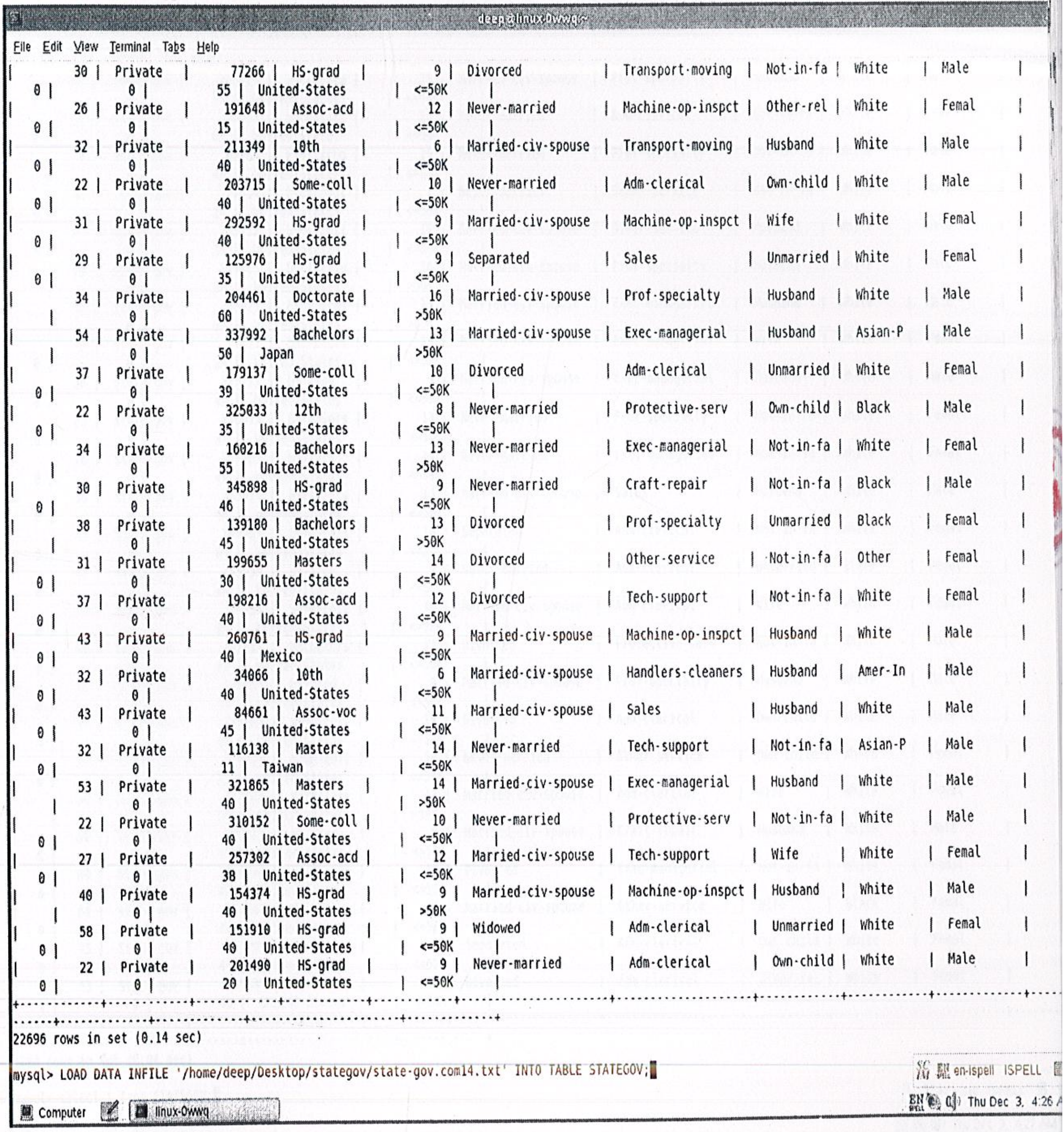
```
35, ?, 35854, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 40, United-States, <=50K
36, ?, 229533, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, <=50K
80, ?, 281768, Assoc-acdm, 12, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 4, United-States, <=50K
29, ?, 499935, Assoc-voc, 11, Never-married, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
18, ?, 200525, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
35, ?, 273558, Some-college, 10, Never-married, ?, Not-in-family, Black, Male, 0, 0, 30, United-States, <=50K
70, ?, 92593, Some-college, 10, Widowed, ?, Not-in-family, White, Female, 0, 0, 25, United-States, <=50K
62, ?, 31577, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 18, United-States, <=50K
18, ?, 90230, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 20, United-States, <=50K
60, ?, 268954, Some-college, 10, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 12, United-States, >50K
52, ?, 89951, 12th, 8, Married-civ-spouse, ?, Wife, Black, Female, 0, 0, 40, United-States, >50K
58, ?, 97969, 1st-4th, 2, Married-spouse-absent, ?, Unmarried, Amer-Indian-Eskimo, Male, 0, 0, 40, United-States, <=50K
62, ?, 178764, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, United-States, >50K
24, ?, 108495, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
26, ?, 375313, Some-college, 10, Never-married, ?, Own-child, Asian-Pac-Islander, Male, 0, 0, 40, Philippines, <=50K
18, ?, 97474, HS-grad, 9, Never-married, ?, Own-child, White, Female, 0, 0, 20, United-States, <=50K
65, ?, 192825, 7th-8th, 4, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
27, ?, 147638, Masters, 14, Never-married, ?, Not-in-family, Other, Female, 0, 0, 40, Japan, <=50K
21, ?, 155697, 9th, 5, Never-married, ?, Own-child, White, Male, 0, 0, 42, United-States, <=50K
51, ?, 43909, HS-grad, 9, Divorced, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 78374, HS-grad, 9, Never-married, ?, Other-relative, Asian-Pac-Islander, Female, 0, 0, 24, United-States, <=50K
62, ?, 263374, Assoc-voc, 11, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 40, Canada, <=50K
34, ?, 330301, 7th-8th, 4, Separated, ?, Unmarried, Black, Female, 0, 0, 40, United-States, <=50K
19, ?, 204868, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 36, United-States, <=50K
49, ?, 113913, HS-grad, 9, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 60, United-States, <=50K
72, ?, 96867, 5th-6th, 3, Widowed, ?, Not-in-family, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 99891, Some-college, 10, Never-married, ?, Own-child, White, Female, 0, 0, 30, United-States, <=50K
59, ?, 120617, Some-college, 10, Never-married, ?, Not-in-family, Black, Female, 0, 0, 40, United-States, <=50K
21, ?, 205939, Some-college, 10, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
63, ?, 126540, Some-college, 10, Divorced, ?, Not-in-family, White, Female, 0, 0, 5, United-States, <=50K
18, ?, 156608, 11th, 7, Never-married, ?, Own-child, White, Female, 0, 0, 25, United-States, <=50K
66, ?, 93318, HS-grad, 9, Widowed, ?, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
20, ?, 203992, HS-grad, 9, Never-married, ?, Own-child, White, Male, 0, 0, 40, United-States, <=50K
49, ?, 114648, 12th, 8, Divorced, ?, Other-relative, Black, Male, 0, 0, 40, United-States, <=50K
60, ?, 134152, 9th, 5, Divorced, ?, Not-in-family, Black, Male, 0, 0, 35, United-States, <=50K
82, ?, 403910, HS-grad, 9, Never-married, ?, Not-in-family, White, Male, 0, 0, 3, United-States, <=50K
81, ?, 120478, Assoc-voc, 11, Divorced, ?, Unmarried, White, Female, 0, 0, 1, ?, <=50K
35, ?, 320084, Bachelors, 13, Married-civ-spouse, ?, Wife, White, Female, 0, 0, 55, United-States, >50K
30, ?, 33811, Bachelors, 13, Never-married, ?, Not-in-family, Asian-Pac-Islander, Female, 0, 0, 99, United-States, <=50K
71, ?, 287372, Doctorate, 16, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 10, United-States, >50K
41, ?, 202822, HS-grad, 9, Separated, ?, Not-in-family, Black, Female, 0, 0, 32, United-States, <=50K
72, ?, 129912, HS-grad, 9, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 25, United-States, <=50K
linux-0wwq:/home/deep/Desktop/unknown # mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.45 SUSE MySQL RPM

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> use adult;
Database changed
mysql> CREATE TABLE UNKNOWN(ATTRIBUTE1 int,ATTRIBUTE2 char(15),ATTRIBUTE3 int,ATTRIBUTE4 varchar(10),ATTRIBUTE5 int,ATTRIBUTE6 char(20),ATTRIBUTE7 char(20),A
TTRIBUTE8 char(10),ATTRIBUTE9 char(8),ATTRIBUTE10 char(6),ATTRIBUTE11 int,ATTRIBUTE12 INT,ATTRIBUTE13 int,ATTRIBUTE14 char(20),ATTRIBUTE15 varchar(6));
```

Private content:



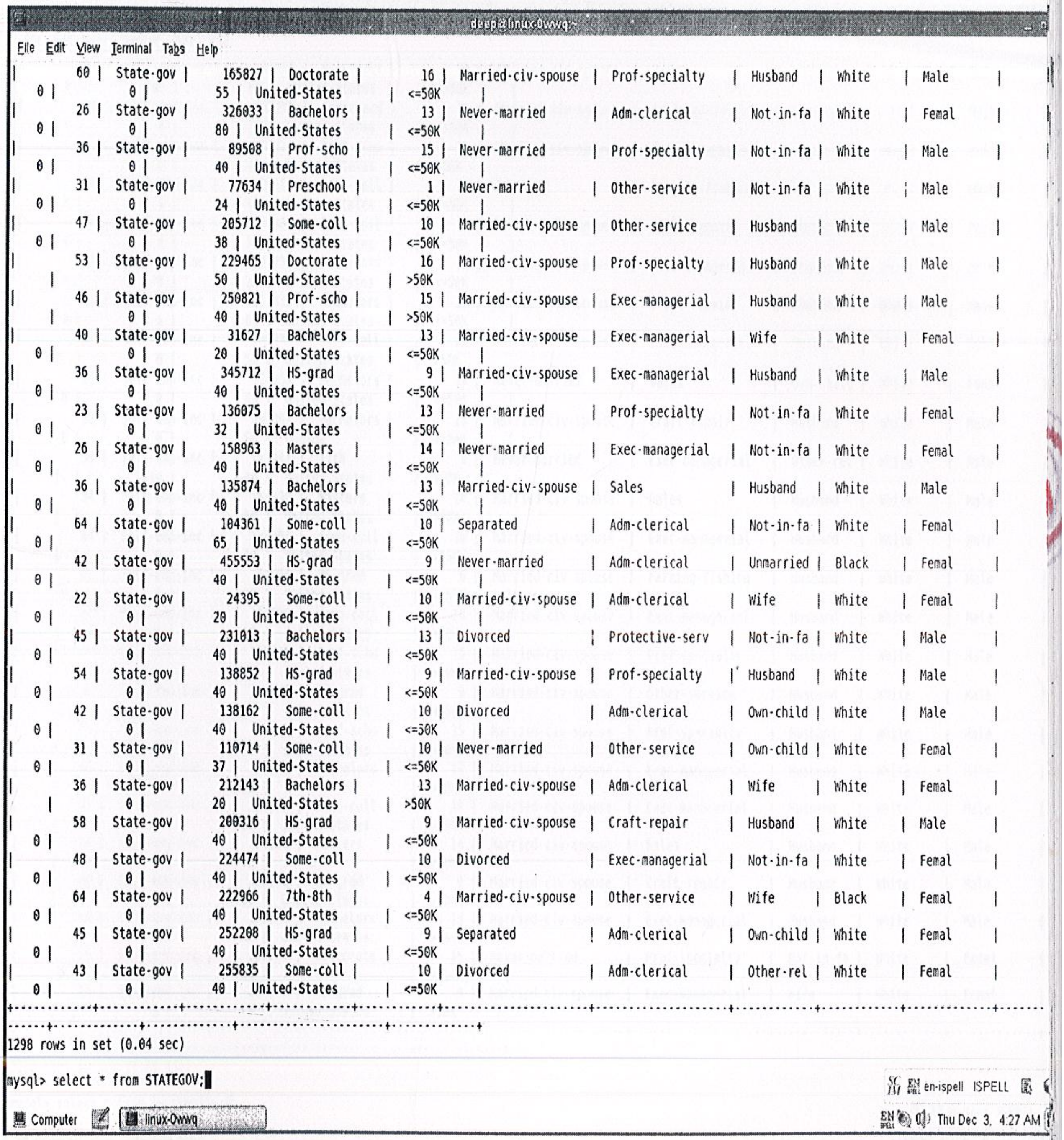
```
|       30 | Private    |      77266 | HS-grad    |          9 | Divorced           | Transport-moving  | Not-in-fa | White     | Male      |
    0 |          0 |         55 | United-States     | <=50K    |
|       26 | Private    |     191648 | Assoc-acd  |         12 | Never-married      | Machine-op-inspct | Other-rel | White     | Femal     |
    0 |          0 |         15 | United-States     | <=50K    |
|       32 | Private    |     211349 | 10th       |          6 | Married-civ-spouse | Transport-moving  | Husband   | White     | Male      |
    0 |          0 |         40 | United-States     | <=50K    |
|       22 | Private    |     203715 | Some-coll  |         10 | Never-married      | Adm-clerical      | Own-child | White     | Male      |
    0 |          0 |         40 | United-States     | <=50K    |
|       31 | Private    |     292592 | HS-grad    |          9 | Married-civ-spouse | Machine-op-inspct | Wife      | White     | Femal     |
    0 |          0 |         40 | United-States     | <=50K    |
|       29 | Private    |     125976 | HS-grad    |          9 | Separated          | Sales             | Unmarried | White     | Femal     |
    0 |          0 |         35 | United-States     | <=50K    |
|       34 | Private    |     204461 | Doctorate  |         16 | Married-civ-spouse | Prof-specialty    | Husband   | White     | Male      |
      |          0 |         60 | United-States     | >50K
|       54 | Private    |     337992 | Bachelors  |         13 | Married-civ-spouse | Exec-managerial   | Husband   | Asian-P   | Male      |
      |          0 |         50 | Japan             | >50K
|       37 | Private    |     179137 | Some-coll  |         10 | Divorced           | Adm-clerical      | Unmarried | White     | Femal     |
    0 |          0 |         39 | United-States     | <=50K    |
|       22 | Private    |     325033 | 12th       |          8 | Never-married      | Protective-serv   | Own-child | Black     | Male      |
    0 |          0 |         35 | United-States     | <=50K    |
|       34 | Private    |     160216 | Bachelors  |         13 | Never-married      | Exec-managerial   | Not-in-fa | White     | Femal     |
      |          0 |         55 | United-States     | >50K
|       30 | Private    |     345898 | HS-grad    |          9 | Never-married      | Craft-repair      | Not-in-fa | Black     | Male      |
    0 |          0 |         46 | United-States     | <=50K    |
|       38 | Private    |     139180 | Bachelors  |         13 | Divorced           | Prof-specialty    | Unmarried | Black     | Femal     |
      |          0 |         45 | United-States     | >50K
|       31 | Private    |     199655 | Masters    |         14 | Divorced           | Other-service     | Not-in-fa | Other     | Femal     |
    0 |          0 |         30 | United-States     | <=50K    |
|       37 | Private    |     198216 | Assoc-acd  |         12 | Divorced           | Tech-support      | Not-in-fa | White     | Femal     |
    0 |          0 |         40 | United-States     | <=50K    |
|       43 | Private    |     260761 | HS-grad    |          9 | Married-civ-spouse | Machine-op-inspct | Husband   | White     | Male      |
    0 |          0 |         40 | Mexico            | <=50K    |
|       32 | Private    |      34066 | 10th       |          6 | Married-civ-spouse | Handlers-cleaners | Husband   | Amer-In   | Male      |
    0 |          0 |         40 | United-States     | <=50K    |
|       43 | Private    |      84661 | Assoc-voc  |         11 | Married-civ-spouse | Sales             | Husband   | White     | Male      |
    0 |          0 |         45 | United-States     | <=50K    |
|       32 | Private    |     116138 | Masters    |         14 | Never-married      | Tech-support      | Not-in-fa | Asian-P   | Male      |
    0 |          0 |         11 | Taiwan            | <=50K    |
|       53 | Private    |     321865 | Masters    |         14 | Married-civ-spouse | Exec-managerial   | Husband   | White     | Male      |
      |          0 |         40 | United-States     | >50K
|       22 | Private    |     310152 | Some-coll  |         10 | Never-married      | Protective-serv   | Not-in-fa | White     | Male      |
    0 |          0 |         40 | United-States     | <=50K    |
|       27 | Private    |     257302 | Assoc-acd  |         12 | Married-civ-spouse | Tech-support      | Wife      | White     | Femal     |
    0 |          0 |         38 | United-States     | <=50K    |
|       40 | Private    |     154374 | HS-grad    |          9 | Married-civ-spouse | Machine-op-inspct | Husband   | White     | Male      |
      |          0 |         40 | United-States     | >50K
|       58 | Private    |     151910 | HS-grad    |          9 | Widowed            | Adm-clerical      | Unmarried | White     | Femal     |
    0 |          0 |         40 | United-States     | <=50K    |
|       22 | Private    |     201490 | HS-grad    |          9 | Never-married      | Adm-clerical      | Own-child | White     | Male      |
    0 |          0 |         20 | United-States     | <=50K    |
+----------+------------+------------+------------+------------+--------------------+-------------------+-----------+-----------+-----------+--------
------+------------+------------+-------------------+----------+
22696 rows in set (0.14 sec)

mysql> select * from PRIVATE;
```

Stategov Data loaded:

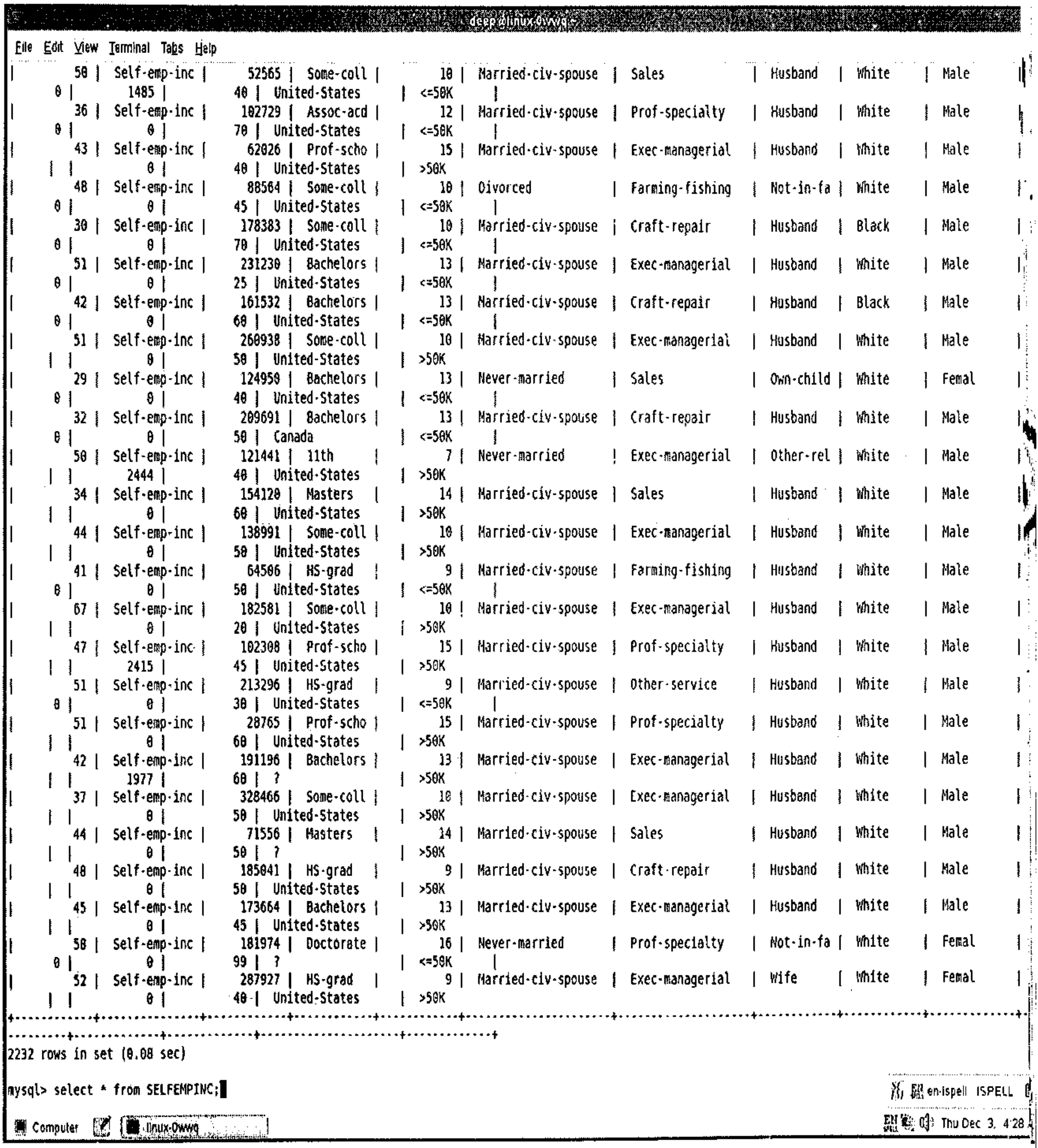


```
|      30 | Private    |     77266 | HS-grad   |         9 | Divorced            | Transport-moving  | Not-in-fa | White     | Male    |
   0 |          0 |         55 | United-States       | <=50K    |
|      26 | Private    |    191648 | Assoc-acd |        12 | Never-married       | Machine-op-inspct | Other-rel | White     | Femal   |
   0 |          0 |         15 | United-States       | <=50K    |
|      32 | Private    |    211349 | 10th      |         6 | Married-civ-spouse  | Transport-moving  | Husband   | White     | Male    |
   0 |          0 |         40 | United-States       | <=50K    |
|      22 | Private    |    203715 | Some-coll |        10 | Never-married       | Adm-clerical      | Own-child | White     | Male    |
   0 |          0 |         40 | United-States       | <=50K    |
|      31 | Private    |    292592 | HS-grad   |         9 | Married-civ-spouse  | Machine-op-inspct | Wife      | White     | Femal   |
   0 |          0 |         40 | United-States       | <=50K    |
|      29 | Private    |    125976 | HS-grad   |         9 | Separated           | Sales             | Unmarried | White     | Femal   |
   0 |          0 |         35 | United-States       | <=50K    |
|      34 | Private    |    204461 | Doctorate |        16 | Married-civ-spouse  | Prof-specialty    | Husband   | White     | Male    |
     |          0 |         60 | United-States       | >50K
|      54 | Private    |    337992 | Bachelors |        13 | Married-civ-spouse  | Exec-managerial   | Husband   | Asian-P   | Male    |
     |          0 |         50 | Japan               | >50K
|      37 | Private    |    179137 | Some-coll |        10 | Divorced            | Adm-clerical      | Unmarried | White     | Femal   |
   0 |          0 |         39 | United-States       | <=50K    |
|      22 | Private    |    325033 | 12th      |         8 | Never-married       | Protective-serv   | Own-child | Black     | Male    |
   0 |          0 |         35 | United-States       | <=50K    |
|      34 | Private    |    160216 | Bachelors |        13 | Never-married       | Exec-managerial   | Not-in-fa | White     | Femal   |
     |          0 |         55 | United-States       | >50K
|      30 | Private    |    345898 | HS-grad   |         9 | Never-married       | Craft-repair      | Not-in-fa | Black     | Male    |
   0 |          0 |         46 | United-States       | <=50K    |
|      38 | Private    |    139180 | Bachelors |        13 | Divorced            | Prof-specialty    | Unmarried | Black     | Femal   |
     |          0 |         45 | United-States       | >50K
|      31 | Private    |    199655 | Masters   |        14 | Divorced            | Other-service     | Not-in-fa | Other     | Femal   |
   0 |          0 |         30 | United-States       | <=50K    |
|      37 | Private    |    198216 | Assoc-acd |        12 | Divorced            | Tech-support      | Not-in-fa | White     | Femal   |
   0 |          0 |         40 | United-States       | <=50K    |
|      43 | Private    |    260761 | HS-grad   |         9 | Married-civ-spouse  | Machine-op-inspct | Husband   | White     | Male    |
   0 |          0 |         40 | Mexico              | <=50K    |
|      32 | Private    |     34066 | 10th      |         6 | Married-civ-spouse  | Handlers-cleaners | Husband   | Amer-In   | Male    |
   0 |          0 |         40 | United-States       | <=50K    |
|      43 | Private    |     84661 | Assoc-voc |        11 | Married-civ-spouse  | Sales             | Husband   | White     | Male    |
   0 |          0 |         45 | United-States       | <=50K    |
|      32 | Private    |    116138 | Masters   |        14 | Never-married       | Tech-support      | Not-in-fa | Asian-P   | Male    |
   0 |          0 |         11 | Taiwan              | <=50K    |
|      53 | Private    |    321865 | Masters   |        14 | Married-civ-spouse  | Exec-managerial   | Husband   | White     | Male    |
     |          0 |         40 | United-States       | >50K
|      22 | Private    |    310152 | Some-coll |        10 | Never-married       | Protective-serv   | Not-in-fa | White     | Male    |
   0 |          0 |         40 | United-States       | <=50K    |
|      27 | Private    |    257302 | Assoc-acd |        12 | Married-civ-spouse  | Tech-support      | Wife      | White     | Femal   |
   0 |          0 |         38 | United-States       | <=50K    |
|      40 | Private    |    154374 | HS-grad   |         9 | Married-civ-spouse  | Machine-op-inspct | Husband   | White     | Male    |
     |          0 |         40 | United-States       | >50K
|      58 | Private    |    151910 | HS-grad   |         9 | Widowed             | Adm-clerical      | Unmarried | White     | Femal   |
   0 |          0 |         40 | United-States       | <=50K    |
|      22 | Private    |    201490 | HS-grad   |         9 | Never-married       | Adm-clerical      | Own-child | White     | Male    |
   0 |          0 |         20 | United-States       | <=50K    |
+-----------+------------+-----------+-----------+-----------+---------------------+-------------------+-----------+-----------+---------+
------+------------+------------+---------------------+----------+
22696 rows in set (0.14 sec)

mysql> LOAD DATA INFILE '/home/deep/Desktop/stategov/state-gov.com14.txt' INTO TABLE STATEGOV;
```

Stategov data selected:

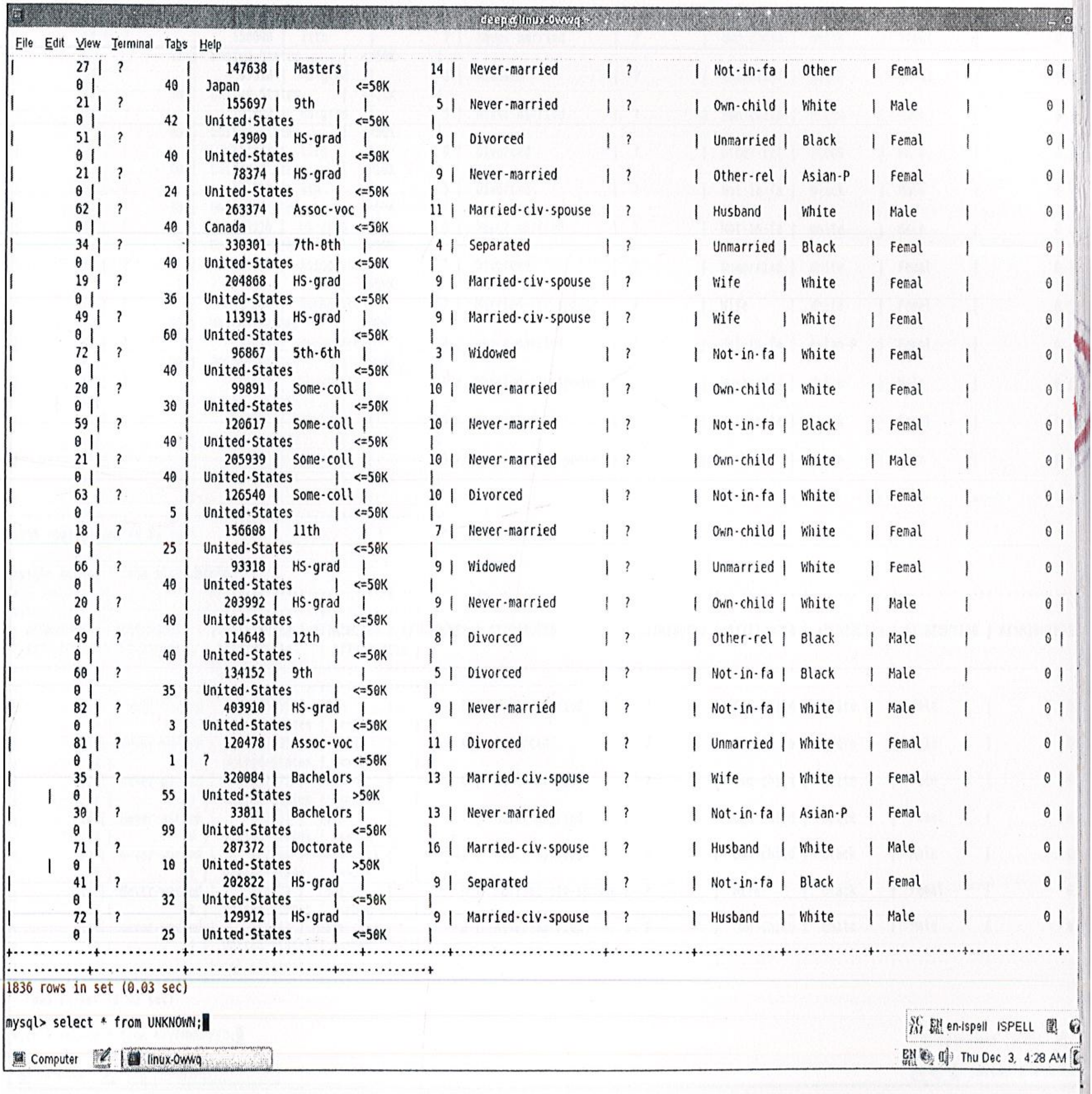


```
|        60 | State-gov |      165827 | Doctorate |          16 | Married-civ-spouse | Prof-specialty   | Husband   | White     | Male      |
    0 |           0 |        55 | United-States   | <=50K |
|        26 | State-gov |      326033 | Bachelors |          13 | Never-married      | Adm-clerical     | Not-in-fa | White     | Femal     |
    0 |           0 |        80 | United-States   | <=50K |
|        36 | State-gov |       89508 | Prof-scho |          15 | Never-married      | Prof-specialty   | Not-in-fa | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        31 | State-gov |       77634 | Preschool |           1 | Never-married      | Other-service    | Not-in-fa | White     | Male      |
    0 |           0 |        24 | United-States   | <=50K |
|        47 | State-gov |      205712 | Some-coll |          10 | Married-civ-spouse | Other-service    | Husband   | White     | Male      |
    0 |           0 |        38 | United-States   | <=50K |
|        53 | State-gov |      229465 | Doctorate |          16 | Married-civ-spouse | Prof-specialty   | Husband   | White     | Male      |
      |           0 |        50 | United-States   | >50K  |
|        46 | State-gov |      250821 | Prof-scho |          15 | Married-civ-spouse | Exec-managerial  | Husband   | White     | Male      |
      |           0 |        40 | United-States   | >50K  |
|        40 | State-gov |       31627 | Bachelors |          13 | Married-civ-spouse | Exec-managerial  | Wife      | White     | Femal     |
    0 |           0 |        20 | United-States   | <=50K |
|        36 | State-gov |      345712 | HS-grad   |           9 | Married-civ-spouse | Exec-managerial  | Husband   | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        23 | State-gov |      136075 | Bachelors |          13 | Never-married      | Prof-specialty   | Not-in-fa | White     | Femal     |
    0 |           0 |        32 | United-States   | <=50K |
|        26 | State-gov |      158963 | Masters   |          14 | Never-married      | Exec-managerial  | Not-in-fa | White     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
|        36 | State-gov |      135874 | Bachelors |          13 | Married-civ-spouse | Sales            | Husband   | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        64 | State-gov |      104361 | Some-coll |          10 | Separated          | Adm-clerical     | Not-in-fa | White     | Femal     |
    0 |           0 |        65 | United-States   | <=50K |
|        42 | State-gov |      455553 | HS-grad   |           9 | Never-married      | Adm-clerical     | Unmarried | Black     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
|        22 | State-gov |       24395 | Some-coll |          10 | Married-civ-spouse | Adm-clerical     | Wife      | White     | Femal     |
    0 |           0 |        20 | United-States   | <=50K |
|        45 | State-gov |      231013 | Bachelors |          13 | Divorced           | Protective-serv  | Not-in-fa | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        54 | State-gov |      138852 | HS-grad   |           9 | Married-civ-spouse | Prof-specialty   | Husband   | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        42 | State-gov |      138162 | Some-coll |          10 | Divorced           | Adm-clerical     | Own-child | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        31 | State-gov |      110714 | Some-coll |          10 | Never-married      | Other-service    | Own-child | White     | Femal     |
    0 |           0 |        37 | United-States   | <=50K |
|        36 | State-gov |      212143 | Bachelors |          13 | Married-civ-spouse | Adm-clerical     | Wife      | White     | Femal     |
      |           0 |        20 | United-States   | >50K  |
|        58 | State-gov |      200316 | HS-grad   |           9 | Married-civ-spouse | Craft-repair     | Husband   | White     | Male      |
    0 |           0 |        40 | United-States   | <=50K |
|        48 | State-gov |      224474 | Some-coll |          10 | Divorced           | Exec-managerial  | Not-in-fa | White     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
|        64 | State-gov |      222966 | 7th-8th   |           4 | Married-civ-spouse | Other-service    | Wife      | Black     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
|        45 | State-gov |      252208 | HS-grad   |           9 | Separated          | Adm-clerical     | Own-child | White     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
|        43 | State-gov |      255835 | Some-coll |          10 | Divorced           | Adm-clerical     | Other-rel | White     | Femal     |
    0 |           0 |        40 | United-States   | <=50K |
+-----------+-----------+-------------+-----------+-------------+--------------------+------------------+-----------+-----------+-----------+--------
------+-------------+-------------+-----------------+-------------+

1298 rows in set (0.04 sec)

mysql> select * from STATEGOV;
```



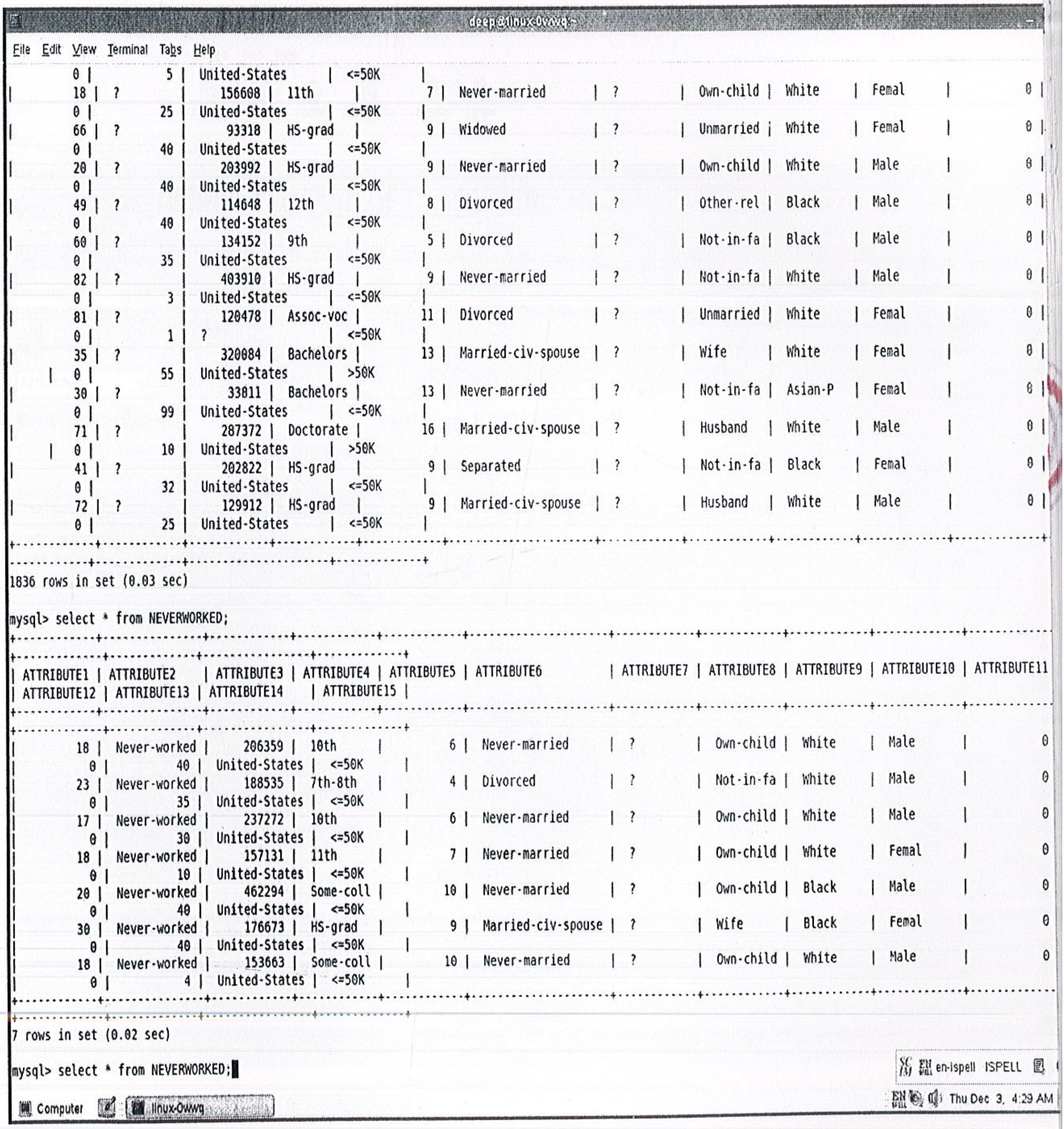
```
|        50 | Self-emp-inc |     52565 | Some-coll |          10 | Married-civ-spouse | Sales           | Husband   | White     | Male     |
     0 |       1485 |          40 | United-States      | <=50K   |
|        36 | Self-emp-inc |    102729 | Assoc-acd |          12 | Married-civ-spouse | Prof-specialty  | Husband   | White     | Male     |
     0 |          0 |          70 | United-States      | <=50K   |
|        43 | Self-emp-inc |     62026 | Prof-scho |          15 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          40 | United-States      | >50K
|        48 | Self-emp-inc |     88564 | Some-coll |          10 | Divorced           | Farming-fishing | Not-in-fa | White     | Male     |
     0 |          0 |          45 | United-States      | <=50K   |
|        30 | Self-emp-inc |    178383 | Some-coll |          10 | Married-civ-spouse | Craft-repair    | Husband   | Black     | Male     |
     0 |          0 |          70 | United-States      | <=50K   |
|        51 | Self-emp-inc |    231230 | Bachelors |          13 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
     0 |          0 |          25 | United-States      | <=50K   |
|        42 | Self-emp-inc |    161532 | Bachelors |          13 | Married-civ-spouse | Craft-repair    | Husband   | Black     | Male     |
     0 |          0 |          60 | United-States      | <=50K   |
|        51 | Self-emp-inc |    260938 | Some-coll |          10 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          50 | United-States      | >50K
|        29 | Self-emp-inc |    124950 | Bachelors |          13 | Never-married      | Sales           | Own-child | White     | Femal    |
     0 |          0 |          40 | United-States      | <=50K   |
|        32 | Self-emp-inc |    209691 | Bachelors |          13 | Married-civ-spouse | Craft-repair    | Husband   | White     | Male     |
     0 |          0 |          50 | Canada             | <=50K   |
|        50 | Self-emp-inc |    121441 | 11th      |           7 | Never-married      | Exec-managerial | Other-rel | White     | Male     |
   | |       2444 |          40 | United-States      | >50K
|        34 | Self-emp-inc |    154120 | Masters   |          14 | Married-civ-spouse | Sales           | Husband   | White     | Male     |
   | |          0 |          60 | United-States      | >50K
|        44 | Self-emp-inc |    138991 | Some-coll |          10 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          50 | United-States      | >50K
|        41 | Self-emp-inc |     64506 | HS-grad   |           9 | Married-civ-spouse | Farming-fishing | Husband   | White     | Male     |
     0 |          0 |          50 | United-States      | <=50K   |
|        67 | Self-emp-inc |    182581 | Some-coll |          10 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          20 | United-States      | >50K
|        47 | Self-emp-inc |    102308 | Prof-scho |          15 | Married-civ-spouse | Prof-specialty  | Husband   | White     | Male     |
   | |       2415 |          45 | United-States      | >50K
|        51 | Self-emp-inc |    213296 | HS-grad   |           9 | Married-civ-spouse | Other-service   | Husband   | White     | Male     |
     0 |          0 |          30 | United-States      | <=50K   |
|        51 | Self-emp-inc |     28765 | Prof-scho |          15 | Married-civ-spouse | Prof-specialty  | Husband   | White     | Male     |
   | |          0 |          60 | United-States      | >50K
|        42 | Self-emp-inc |    191196 | Bachelors |          13 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |       1977 |          60 | ?                  | >50K
|        37 | Self-emp-inc |    328466 | Some-coll |          10 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          50 | United-States      | >50K
|        44 | Self-emp-inc |     71556 | Masters   |          14 | Married-civ-spouse | Sales           | Husband   | White     | Male     |
   | |          0 |          50 | ?                  | >50K
|        40 | Self-emp-inc |    185041 | HS-grad   |           9 | Married-civ-spouse | Craft-repair    | Husband   | White     | Male     |
   | |          0 |          50 | United-States      | >50K
|        45 | Self-emp-inc |    173664 | Bachelors |          13 | Married-civ-spouse | Exec-managerial | Husband   | White     | Male     |
   | |          0 |          45 | United-States      | >50K
|        58 | Self-emp-inc |    181974 | Doctorate |          16 | Never-married      | Prof-specialty  | Not-in-fa | White     | Femal    |
     0 |          0 |          99 | ?                  | <=50K   |
|        52 | Self-emp-inc |    287927 | HS-grad   |           9 | Married-civ-spouse | Exec-managerial | Wife      | White     | Femal    |
   | |          0 |          40 | United-States      | >50K
+-----------+--------------+-----------+-----------+-------------+--------------------+-----------------+-----------+-----------+----------+
---------+-------------+-------------+--------------------+-------------+
2232 rows in set (0.08 sec)

mysql> select * from SELFEMPINC;
```

Unknown Content:



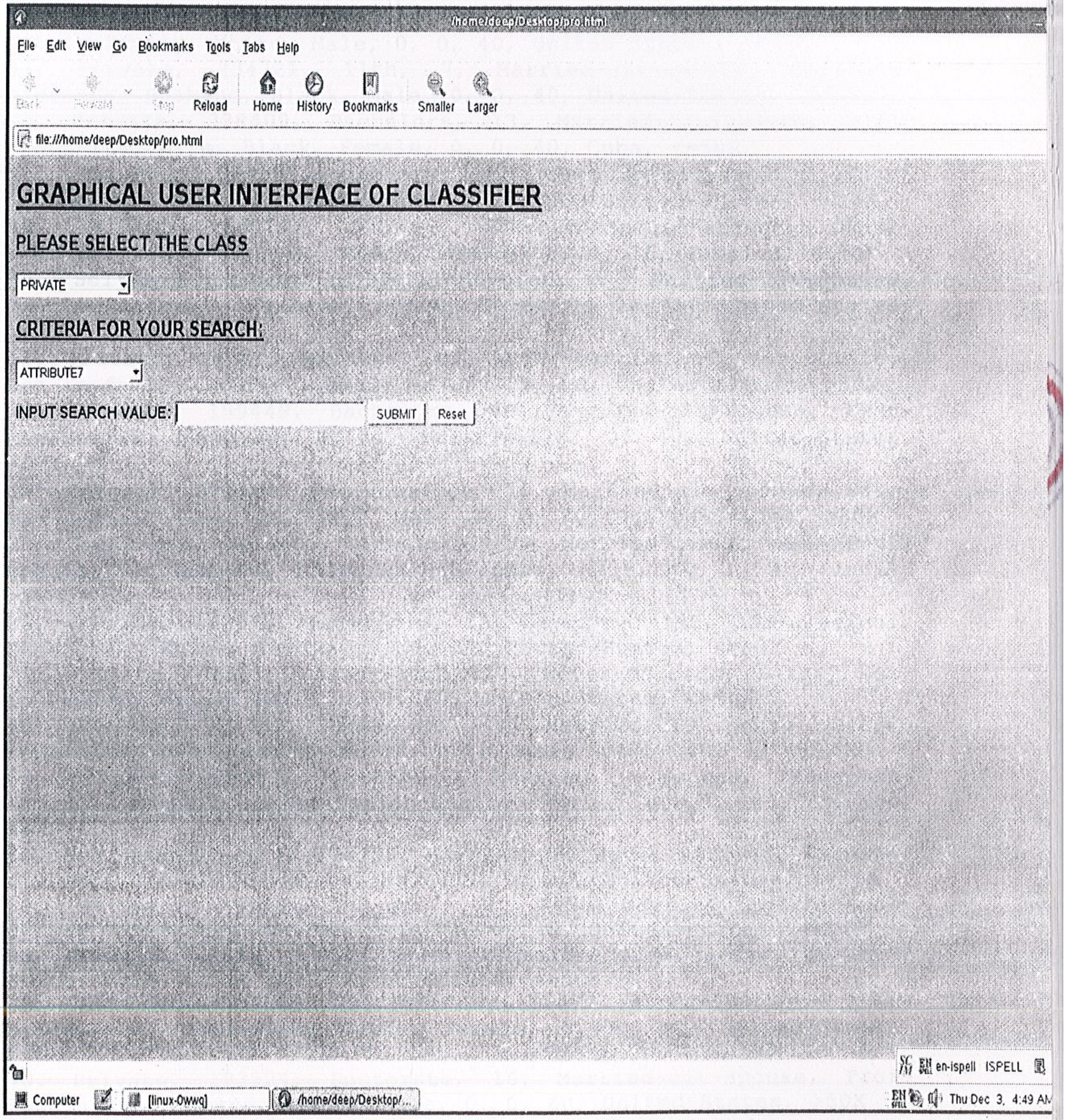
```
|          27 | ?          |      147638 | Masters    |          14 | Never-married      | ?          | Not-in-fa | Other     | Femal      |            0 |
            0 |         40 | Japan              | <=50K       |
|          21 | ?          |      155697 | 9th        |           5 | Never-married      | ?          | Own-child | White     | Male       |            0 |
            0 |         42 | United-States      | <=50K       |
|          51 | ?          |       43909 | HS-grad    |           9 | Divorced           | ?          | Unmarried | Black     | Femal      |            0 |
            0 |         40 | United-States      | <=50K       |
|          21 | ?          |       78374 | HS-grad    |           9 | Never-married      | ?          | Other-rel | Asian-P   | Femal      |            0 |
            0 |         24 | United-States      | <=50K       |
|          62 | ?          |      263374 | Assoc-voc  |          11 | Married-civ-spouse | ?          | Husband   | White     | Male       |            0 |
            0 |         40 | Canada             | <=50K       |
|          34 | ?          |      330301 | 7th-8th    |           4 | Separated          | ?          | Unmarried | Black     | Femal      |            0 |
            0 |         40 | United-States      | <=50K       |
|          19 | ?          |      204868 | HS-grad    |           9 | Married-civ-spouse | ?          | Wife      | White     | Femal      |            0 |
            0 |         36 | United-States      | <=50K       |
|          49 | ?          |      113913 | HS-grad    |           9 | Married-civ-spouse | ?          | Wife      | White     | Femal      |            0 |
            0 |         60 | United-States      | <=50K       |
|          72 | ?          |       96867 | 5th-6th    |           3 | Widowed            | ?          | Not-in-fa | White     | Femal      |            0 |
            0 |         40 | United-States      | <=50K       |
|          20 | ?          |       99891 | Some-coll  |          10 | Never-married      | ?          | Own-child | White     | Femal      |            0 |
            0 |         30 | United-States      | <=50K       |
|          59 | ?          |      120617 | Some-coll  |          10 | Never-married      | ?          | Not-in-fa | Black     | Femal      |            0 |
            0 |         40 | United-States      | <=50K       |
|          21 | ?          |      205939 | Some-coll  |          10 | Never-married      | ?          | Own-child | White     | Male       |            0 |
            0 |         40 | United-States      | <=50K       |
|          63 | ?          |      126540 | Some-coll  |          10 | Divorced           | ?          | Not-in-fa | White     | Femal      |            0 |
            0 |          5 | United-States      | <=50K       |
|          18 | ?          |      156608 | 11th       |           7 | Never-married      | ?          | Own-child | White     | Femal      |            0 |
            0 |         25 | United-States      | <=50K       |
|          66 | ?          |       93318 | HS-grad    |           9 | Widowed            | ?          | Unmarried | White     | Femal      |            0 |
            0 |         40 | United-States      | <=50K       |
|          20 | ?          |      203992 | HS-grad    |           9 | Never-married      | ?          | Own-child | White     | Male       |            0 |
            0 |         40 | United-States      | <=50K       |
|          49 | ?          |      114648 | 12th       |           8 | Divorced           | ?          | Other-rel | Black     | Male       |            0 |
            0 |         40 | United-States      | <=50K       |
|          60 | ?          |      134152 | 9th        |           5 | Divorced           | ?          | Not-in-fa | Black     | Male       |            0 |
            0 |         35 | United-States      | <=50K       |
|          82 | ?          |      403910 | HS-grad    |           9 | Never-married      | ?          | Not-in-fa | White     | Male       |            0 |
            0 |          3 | United-States      | <=50K       |
|          81 | ?          |      120478 | Assoc-voc  |          11 | Divorced           | ?          | Unmarried | White     | Femal      |            0 |
            0 |          1 | ?                  | <=50K       |
|          35 | ?          |      320084 | Bachelors  |          13 | Married-civ-spouse | ?          | Wife      | White     | Femal      |            0 |
      |     0 |         55 | United-States      | >50K
|          30 | ?          |       33811 | Bachelors  |          13 | Never-married      | ?          | Not-in-fa | Asian-P   | Femal      |            0 |
            0 |         99 | United-States      | <=50K       |
|          71 | ?          |      287372 | Doctorate  |          16 | Married-civ-spouse | ?          | Husband   | White     | Male       |            0 |
      |     0 |         10 | United-States      | >50K
|          41 | ?          |      202822 | HS-grad    |           9 | Separated          | ?          | Not-in-fa | Black     | Femal      |            0 |
            0 |         32 | United-States      | <=50K       |
|          72 | ?          |      129912 | HS-grad    |           9 | Married-civ-spouse | ?          | Husband   | White     | Male       |            0 |
            0 |         25 | United-States      | <=50K       |
+-------------+------------+-------------+------------+-------------+--------------------+------------+-----------+-----------+------------+--------------+
-------------+------------+--------------------+-------------+
1836 rows in set (0.03 sec)

mysql> select * from UNKNOWN;
```

Networked Content:



```
          0 |            5 | United-States      | <=50K      |
|        18 | ?            |     156608 | 11th         |          7 | Never-married        | ?            | Own-child | White      | Femal      |           0 |
          0 |           25 | United-States      | <=50K      |
|        66 | ?            |      93318 | HS-grad      |          9 | Widowed              | ?            | Unmarried | White      | Femal      |           0 |
          0 |           40 | United-States      | <=50K      |
|        20 | ?            |     203992 | HS-grad      |          9 | Never-married        | ?            | Own-child | White      | Male       |           0 |
          0 |           40 | United-States      | <=50K      |
|        49 | ?            |     114648 | 12th         |          8 | Divorced             | ?            | Other-rel | Black      | Male       |           0 |
          0 |           40 | United-States      | <=50K      |
|        60 | ?            |     134152 | 9th          |          5 | Divorced             | ?            | Not-in-fa | Black      | Male       |           0 |
          0 |           35 | United-States      | <=50K      |
|        82 | ?            |     403910 | HS-grad      |          9 | Never-married        | ?            | Not-in-fa | White      | Male       |           0 |
          0 |            3 | United-States      | <=50K      |
|        81 | ?            |     120478 | Assoc-voc    |         11 | Divorced             | ?            | Unmarried | White      | Femal      |           0 |
          0 |            1 | ?                  | <=50K      |
|        35 | ?            |     320084 | Bachelors    |         13 | Married-civ-spouse   | ?            | Wife      | White      | Femal      |           0 |
      |   0 |           55 | United-States      | >50K
|        30 | ?            |      33811 | Bachelors    |         13 | Never-married        | ?            | Not-in-fa | Asian-P    | Femal      |           0 |
          0 |           99 | United-States      | <=50K      |
|        71 | ?            |     287372 | Doctorate    |         16 | Married-civ-spouse   | ?            | Husband   | White      | Male       |           0 |
      |   0 |           10 | United-States      | >50K
|        41 | ?            |     202822 | HS-grad      |          9 | Separated            | ?            | Not-in-fa | Black      | Femal      |           0 |
          0 |           32 | United-States      | <=50K      |
|        72 | ?            |     129912 | HS-grad      |          9 | Married-civ-spouse   | ?            | Husband   | White      | Male       |           0 |
          0 |           25 | United-States      | <=50K      |
+-----------+--------------+--------------------+------------+------------+---------------------+--------------+-----------+------------+------------+-------------+
------------+--------------+--------------------+------------+
1836 rows in set (0.03 sec)

mysql> select * from NEVERWORKED;
+------------+--------------+-------------+------------+-------------+-------------------+-------------+-------------+-------------+--------------+------------
+------------+--------------+-------------+-------------+
| ATTRIBUTE1 | ATTRIBUTE2   | ATTRIBUTE3 | ATTRIBUTE4 | ATTRIBUTE5 | ATTRIBUTE6          | ATTRIBUTE7 | ATTRIBUTE8 | ATTRIBUTE9 | ATTRIBUTE10 | ATTRIBUTE11
| ATTRIBUTE12 | ATTRIBUTE13 | ATTRIBUTE14    | ATTRIBUTE15 |
+------------+--------------+-------------+------------+-------------+-------------------+-------------+-------------+-------------+--------------+------------
+------------+--------------+-------------+-------------+
|         18 | Never-worked |      206359 | 10th       |           6 | Never-married      | ?           | Own-child | White       | Male         |           0
|          0 |           40 | United-States | <=50K       |
|         23 | Never-worked |      188535 | 7th-8th    |           4 | Divorced           | ?           | Not-in-fa | White       | Male         |           0
|          0 |           35 | United-States | <=50K       |
|         17 | Never-worked |      237272 | 10th       |           6 | Never-married      | ?           | Own-child | White       | Male         |           0
|          0 |           30 | United-States | <=50K       |
|         18 | Never-worked |      157131 | 11th       |           7 | Never-married      | ?           | Own-child | White       | Femal        |           0
|          0 |           10 | United-States | <=50K       |
|         20 | Never-worked |      462294 | Some-coll  |          10 | Never-married      | ?           | Own-child | Black       | Male         |           0
|          0 |           40 | United-States | <=50K       |
|         30 | Never-worked |      176673 | HS-grad    |           9 | Married-civ-spouse | ?           | Wife      | Black       | Femal        |           0
|          0 |           40 | United-States | <=50K       |
|         18 | Never-worked |      153663 | Some-coll  |          10 | Never-married      | ?           | Own-child | White       | Male         |           0
|          0 |            4 | United-States | <=50K       |
+------------+--------------+-------------+------------+-------------+-------------------+-------------+-------------+-------------+--------------+------------
+------------+--------------+-------------+-------------+
7 rows in set (0.02 sec)

mysql> select * from NEVERWORKED;
```

Classifier:


/home/deep/Desktop/pro.html
File Edit View Go Bookmarks Tools Tabs Help
Back Forward Stop Reload Home History Bookmarks Smaller Larger
file:///home/deep/Desktop/pro.html
SUBMIT
Reset
en-ispell ISPELL
Computer
[linux-Owwq]
/home/deep/Desktop/...
Thu Dec 3, 4:49 AM

GRAPHICAL USER INTERFACE OF CLASSIFIER

PLEASE SELECT THE CLASS

PRIVATE

CRITERIA FOR YOUR SEARCH:

ATTRIBUTE7

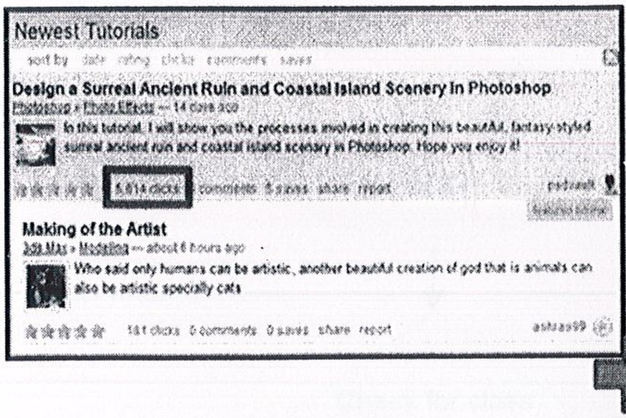
INPUT SEARCH VALUE:

Sample Adult Data:

```
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical,
Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse,
Exec-managerial, Husband, White, Male, 0, 0, 13, United-States,
<=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners,
Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-
cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-
specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
37, Private, 284582, Masters, 14, Married-civ-spouse, Exec-
managerial, Wife, White, Female, 0, 0, 40, United-States, <=50K
49, Private, 160187, 9th, 5, Married-spouse-absent, Other-
service, Not-in-family, Black, Female, 0, 0, 16, Jamaica, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse,
Exec-managerial, Husband, White, Male, 0, 0, 45, United-States,
>50K
31, Private, 45781, Masters, 14, Never-married, Prof-specialty,
Not-in-family, White, Female, 14084, 0, 50, United-States, >50K
42, Private, 159449, Bachelors, 13, Married-civ-spouse, Exec-
managerial, Husband, White, Male, 5178, 0, 40, United-States,
>50K
37, Private, 280464, Some-college, 10, Married-civ-spouse, Exec-
managerial, Husband, Black, Male, 0, 0, 80, United-States, >50K
30, State-gov, 141297, Bachelors, 13, Married-civ-spouse, Prof-
specialty, Husband, Asian-Pac-Islander, Male, 0, 0, 40, India,
>50K
23, Private, 122272, Bachelors, 13, Never-married, Adm-clerical,
Own-child, White, Female, 0, 0, 30, United-States, <=50K
32, Private, 205019, Assoc-acdm, 12, Never-married, Sales, Not-
in-family, Black, Male, 0, 0, 50, United-States, <=50K
40, Private, 121772, Assoc-voc, 11, Married-civ-spouse, Craft-
repair, Husband, Asian-Pac-Islander, Male, 0, 0, 40, ?, >50K
34, Private, 245487, 7th-8th, 4, Married-civ-spouse, Transport-
moving, Husband, Amer-Indian-Eskimo, Male, 0, 0, 45, Mexico,
<=50K
25, Self-emp-not-inc, 176756, HS-grad, 9, Never-married, Farming-
fishing, Own-child, White, Male, 0, 0, 35, United-States, <=50K
32, Private, 186824, HS-grad, 9, Never-married, Machine-op-
inspct, Unmarried, White, Male, 0, 0, 40, United-States, <=50K
38, Private, 28887, 11th, 7, Married-civ-spouse, Sales, Husband,
White, Male, 0, 0, 50, United-States, <=50K
43, Self-emp-not-inc, 292175, Masters, 14, Divorced, Exec-
managerial, Unmarried, White, Female, 0, 0, 45, United-States,
>50K
40, Private, 193524, Doctorate, 16, Married-civ-spouse, Prof-
specialty, Husband, White, Male, 0, 0, 60, United-States, >50K
54, Private, 302146, HS-grad, 9, Separated, Other-service,
Unmarried, Black, Female, 0, 0, 20, United-States, <=50K
```

```
35, Federal-gov, 76845, 9th, 5, Married-civ-spouse, Farming-fishing, Husband, Black, Male, 0, 0, 40, United-States, <=50K
43, Private, 117037, 11th, 7, Married-civ-spouse, Transport-moving, Husband, White, Male, 0, 2042, 40, United-States, <=50K
59, Private, 109015, HS-grad, 9, Divorced, Tech-support, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
56, Local-gov, 216851, Bachelors, 13, Married-civ-spouse, Tech-support, Husband, White, Male, 0, 0, 40, United-States, >50K
19, Private, 168294, HS-grad, 9, Never-married, Craft-repair, Own-child, White, Male, 0, 0, 40, United-States, <=50K
54, ?, 180211, Some-college, 10, Married-civ-spouse, ?, Husband, Asian-Pac-Islander, Male, 0, 0, 60, South, >50K
39, Private, 367260, HS-grad, 9, Divorced, Exec-managerial, Not-in-family, White, Male, 0, 0, 80, United-States, <=50K
49, Private, 193366, HS-grad, 9, Married-civ-spouse, Craft-repair, Husband, White, Male, 0, 0, 40, United-States, <=50K
23, Local-gov, 190709, Assoc-acdm, 12, Never-married, Protective-serv, Not-in-family, White, Male, 0, 0, 52, United-States, <=50K
20, Private, 266015, Some-college, 10, Never-married, Sales, Own-child, Black, Male, 0, 0, 44, United-States, <=50K
45, Private, 386940, Bachelors, 13, Divorced, Exec-managerial, Own-child, White, Male, 0, 1408, 40, United-States, <=50K
30, Federal-gov, 59951, Some-college, 10, Married-civ-spouse, Adm-clerical, Own-child, White, Male, 0, 0, 40, United-States, <=50K
22, State-gov, 311512, Some-college, 10, Married-civ-spouse, Other-service, Husband, Black, Male, 0, 0, 15, United-States, <=50K
48, Private, 242406, 11th, 7, Never-married, Machine-op-inspct, Unmarried, White, Male, 0, 0, 40, Puerto-Rico, <=50K
21, Private, 197200, Some-college, 10, Never-married, Machine-op-inspct, Own-child, White, Male, 0, 0, 40, United-States, <=50K
19, Private, 544091, HS-grad, 9, Married-AF-spouse, Adm-clerical, Wife, White, Female, 0, 0, 25, United-States, <=50K
31, Private, 84154, Some-college, 10, Married-civ-spouse, Sales, Husband, White, Male, 0, 0, 38, ?, >50K
48, Self-emp-not-inc, 265477, Assoc-acdm, 12, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 40, United-States, <=50K
31, Private, 507875, 9th, 5, Married-civ-spouse, Machine-op-inspct, Husband, White, Male, 0, 0, 43, United-States, <=50K
53, Self-emp-not-inc, 88506, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, White, Male, 0, 0, 40, United-States, <=50K
24, Private, 172987, Bachelors, 13, Married-civ-spouse, Tech-support, Husband, White, Male, 0, 0, 50, United-States, <=50K
49, Private, 94638, HS-grad, 9, Separated, Adm-clerical, Unmarried, White, Female, 0, 0, 40, United-States, <=50K
25, Private, 289980, HS-grad, 9, Never-married, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 35, United-States, <=50K
```

57, Federal-gov, 337895, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, Black, Male, 0, 0, 40, United-States, >50K



Newest Tutorials

Design a Surreal Ancient Ruin and Coastal Island Scenery in Photoshop

Photoshop » Photo Effects — 14 days ago

In this tutorial, I will show you the processes involved in creating this beautiful, fantasy-styled surreal ancient ruin and coastal island scenary in Photoshop. Hope you enjoy it!

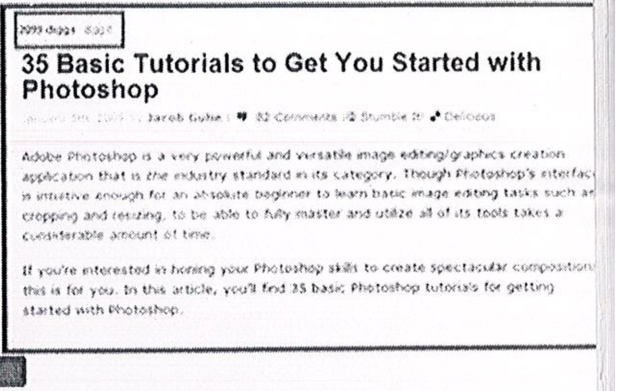
6,614 clicks 0 comments 5 saves share report psdvault

Making of the Artist

3ds Max » Modeling — about 6 hours ago

Who said only humans can be artistic, another beautiful creation of god that is animals can also be artistic specially cats

181 clicks 0 comments 0 saves share report ashraa99



2099 diggs digg it

35 Basic Tutorials to Get You Started with Photoshop

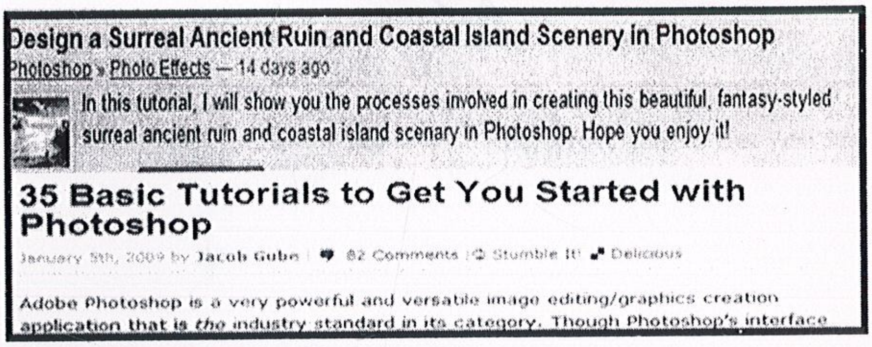
January 5th, 2009 by Jacob Gube | 82 Comments | Stumble It! Delicious

Adobe Photoshop is a very powerful and versatile image editing/graphics creation application that is *the* industry standard in its category. Though Photoshop's interfac is intuitive enough for an absolute beginner to learn basic image editing tasks such as cropping and resizing, to be able to fully master and utilize all of its tools takes a considerable amount of time.

If you're interested in honing your Photoshop skills to create spectacular composition this is for you. In this article, you'll find 35 basic Photoshop tutorials for getting started with Photoshop.

+

=



Design a Surreal Ancient Ruin and Coastal Island Scenery in Photoshop

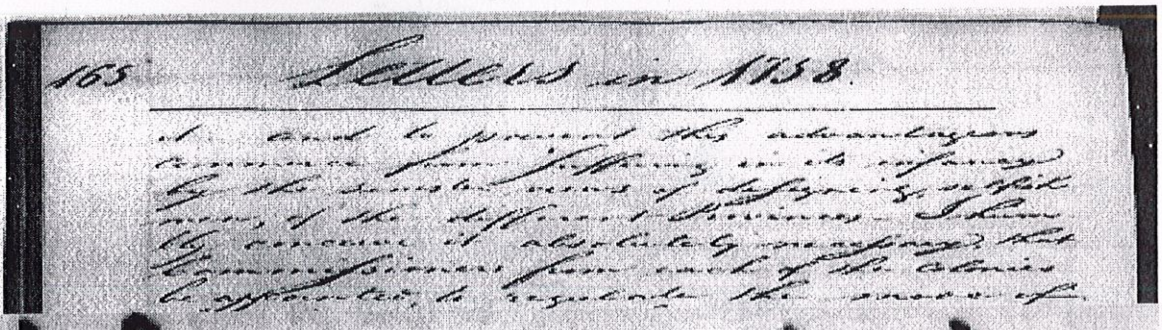
Photoshop » Photo Effects — 14 days ago

In this tutorial, I will show you the processes involved in creating this beautiful, fantasy-styled surreal ancient ruin and coastal island scenary in Photoshop. Hope you enjoy it!

35 Basic Tutorials to Get You Started with Photoshop

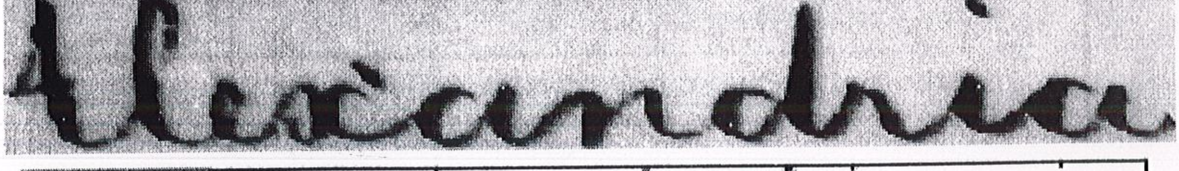
January 5th, 2009 by Jacob Gube | 82 Comments | Stumble It! Delicious

Adobe Photoshop is a very powerful and versatile image editing/graphics creation application that is *the* industry standard in its category. Though Photoshop's interface



165 Letters in 1838.

[illegible]



Alexandria

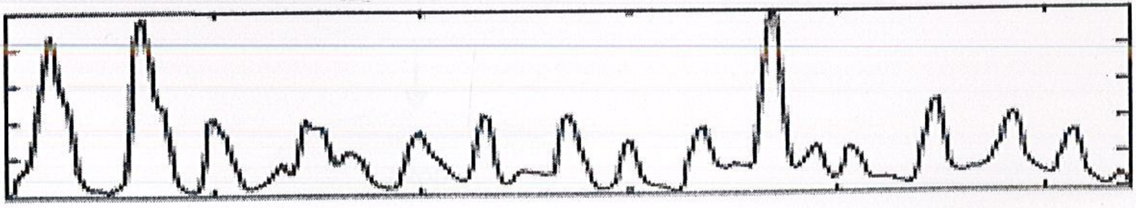
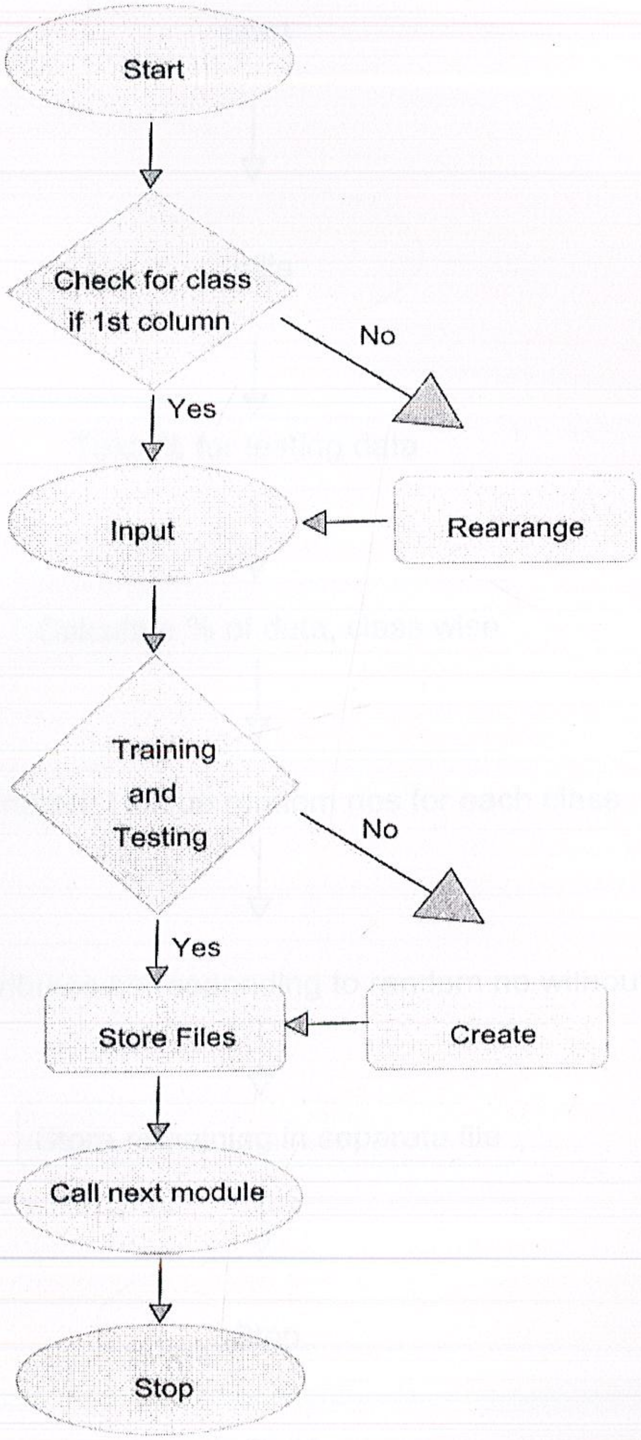


Fig 8.8



Start
Check for class if 1st column
No
Yes
Input
Rearrange
Training and Testing
No
Yes
Store Files
Create
Call next module
Stop

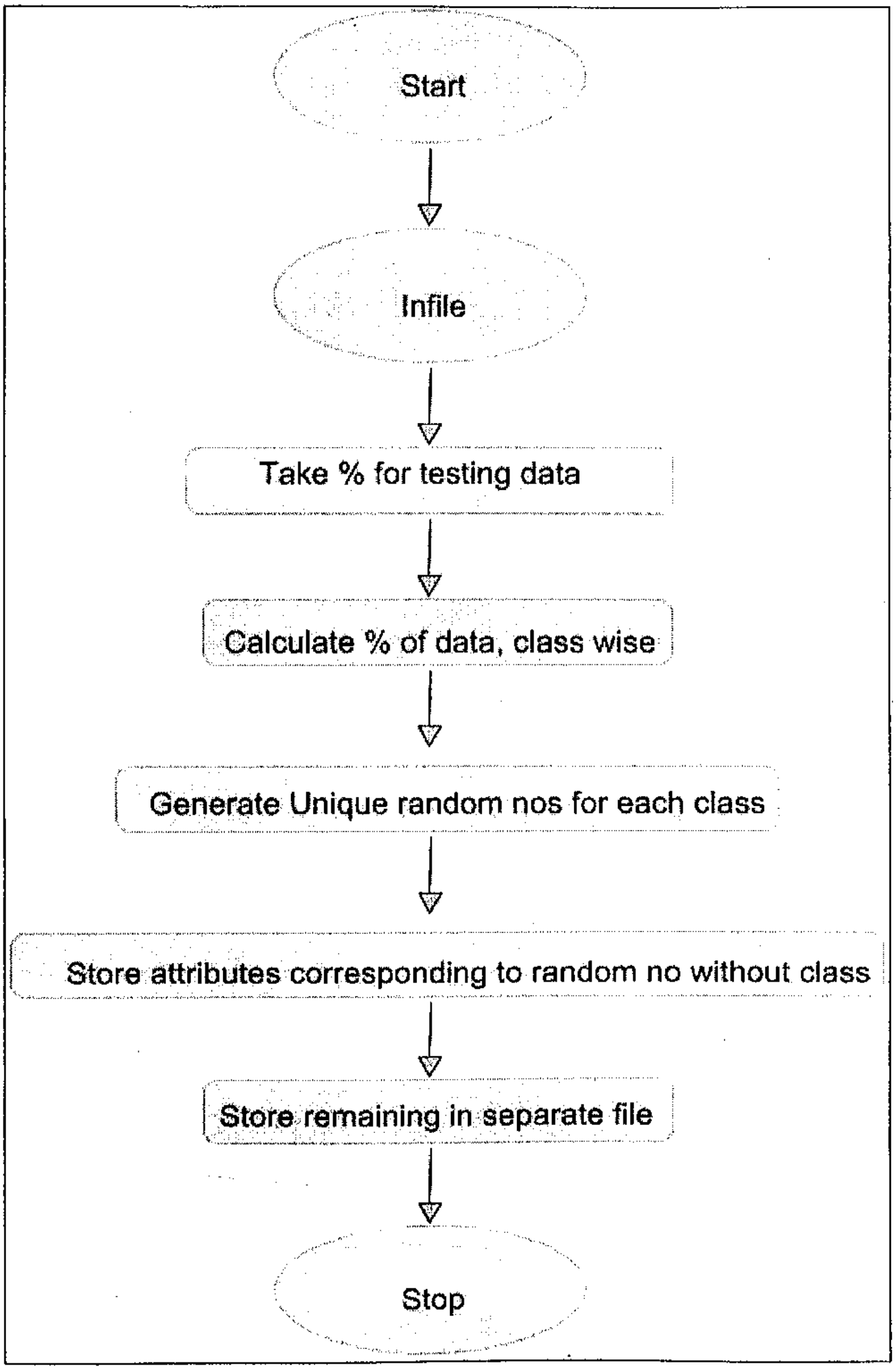

Start
Infile
Take % for testing data
Calculate % of data, class wise
Generate Unique random nos for each class
Store attributes corresponding to random no without class
Store remaining in separate file
Stop

Fig 8.9

CHAPTER 9

SAMPLE CODES

9.1 Sample code for generating training and testing datasets

```
for(i=0;i<=maxclass;i++)
{

    if(count[i]==0)
        continue;
        else
        {

        for(j=0;j<count_test[i];j++)
        {

         rand_no[j]=rand()%count[i];

        int spam=0;
        while(spam==0)
        {

        for(l=0;l<j;l++)
        {

        if(rand_no[j]==rand_no[l])
        {

          rand_no[j]=rand()%count[i];
           l=0;
        }
        }

        if(j==l)
        {
         spam=1;
         }
        }

                        //          printf("\nRandom number generated is   -->
%d",rand_no[j]);
```

```
//                              printf("\n Random no gen class %d --> %d",i,rand_no[j]);
                                b=0;
                                for(k=1;k<maxc;k++)
                                {

                                fprintf(out_test,"%3.2f ",arry[i][rand_no[j]][k]);
                                }
                                for(b=0;b<maxc;b++)
                                {

                                        arrtst[a][b]=arry[i][rand_no[j]][b+1];
                                //      printf("%3.2f ",arrtst[a][b]);

                                }

                                arrtst1[a]=arry[i][rand_no[j]][0];
//                              printf("\n %1.0f--%1.0f--
%d",arrtst1[a],arry[i][rand_no[j]][0],a);
                                a++;
//                              printf("\n");
                                fprintf(out_test,"\n");

                        }

//                      printf("\nPrinting the training data-->");

                        for(m=0;m<count[i];m++)

                        {//

                                printf("\n %d",m);
//                              printf("\nDone i");
                                done=0;

                                for(n=0;n<count_test[i];n++)

                                {
//
                                        printf("\t%d   %d",m,rand_no[n]);
                                        if(m==rand_no[n])

                                        {

                                                done=1;
```

```
                    }

                }
                if(done==0)
                {

//                  printf("\nPrinting the array");
                    for(k=0;k<maxc;k++)
                    {

                    fprintf(out_train,"%3.2f ",arry[i][m][k]);

                    }
//                  printf("\n<%d>",c);
                    for(k=0;k<maxc;k++)
                    {

                        arrtr[c][k]=arry[i][m][k];
//                      printf("%3.2f ",arrtr[c][k]);

                    }
                    c++;
                    if(c>=train_count)
                    break;
                    fprintf(out_train,"\n");
                }

//              printf("\n%d",c);

            }
//      c++;

        }

        printf(" ");
    }
```

9.2 Sample code for distance calculation

```
void eucleid()
{

        sl=(int*)malloc(sizeof(int)*train_count);
        wt=(int*)malloc(sizeof(int)*kn);
        wtlb=(int*)malloc(sizeof(int)*kn);
        dist=(float**)malloc(sizeof(float)*train_count);
        for(i=0;i<train_count;i++)
        dist[i]=(float*)malloc(sizeof(float)*maxc-1);
        tdist=(float*)malloc(sizeof(float)*train_count);
        for(i=0;i<test_count;i++)
        {

//              printf("\nDistance '%d'th test data  -->  ",i+1);
                for(j=0;j<train_count;j++)
                {

                        for(k=0;k<maxc-1;k++)
                        {

                                dist[j][k]=pow((arrtst[i][k]-arrtr[j][k+1]),2);
                                tot_dist=tot_dist+dist[j][k];
//                              printf("\n%3.2f ",tot_dist);
//                              printf("-<%3.2f %3.2f>-",arrtst[i][k],arrtr[j][k]);

        }

                        tdist[j]=sqrt(tot_dist);
                        sl[j]=j;
//                      printf("%3.2f\t",tdist[j]);
                        tot_dist=0;
                }

                for (j = (train_count - 1); j >= 0; j--)
                {

                        for (k = 1; k <= j; k++)
                        {

                                if (tdist[k-1] > tdist[k])
                                {
                                        temp = tdist[k-1];
                                        tdist[k-1] = tdist[k];
                                        tdist[k] = temp;
                                        tempsl=sl[k-1];
                                        sl[k-1]=sl[k];
                                        sl[k]=tempsl;

                                }

                        }

                }

/*              for(j=0;j<train_count;j++)
                {

                        printf("%d-",sl[j]);
                }

                for(j=0;j<train_count;j++)
                {

                        printf("%1.0f ",arrtr[j][0]);

                }*/
                for(j=0;j<kn;j++)
                {

//                      printf("%3.2f<%1.0f>  ",tdist[j],arrtr[sl[j]][0]);
                }
//              printf("\n");
                for(k=0;k<kn;k++)
                {

                        wt[k]=kn-k;
                        c=arrtr[sl[k]][0];
                        wtlb[c]+=wt[k];
//                      printf("%d ",wtlb[c]);
                }

/*              for(k=0;k<=maxclass;k++)
                {

                        c=arrtr[sl[k]][0];
                        printf("wtlb[%d]->%d ",k,wtlb[k]);
```

```
            }*/

            max=wtlb[0];
            c=0;
            for(k=1;k<=maxclass;k++)
            {

                    if(wtlb[k]>wtlb[k-1])
                    {

                            max=wtlb[k];
                            c=k;

            }
                    else

                    {

                            max=max;
                            c=c;
                    }

//                  printf("%d-",c);

            }

//          printf("\nPredictedclass is -->  %d",c);
            float ct=c;
            if(ct==arrtst1[i])
            {

//          printf("\nPrediction Correct  ");
            countot++;

            }
            else
            {

//          printf("\nPrediction Incorrect");
            printf(" ");
            }

/*          int yui;
            int countot=0;
            int kuchbhi;
            kuchbhi=arrtst[i][0];
```

```
        for(k=0;k<maxclass;k++)
        {

        if(count_test[k]!=0)
        {

        countot=countot+count_test[k]-1;

        }
                if(i>=countot && i<=count_test[k]+count_test[k+1])
                {

                        if(c==kuchbhi)
                        {

                                printf("\nPrediction correct");
                                corrcount++;

                        }
                        else
                        {

                                printf("\nPrediction is incorrect");
                        }

                }
                else
                {

                        continue;
                }

        }*/
        for(k=0;k<kn;k++)
        {

                wt[k]=0;
        }
        for(k=0;k<=maxclass;k++)
        {

                wtlb[k]=0;
        }
}

float smthing;
```

```
        printf("%d--%d",countot,test_count);
        smthing=(countot*100)/test_count;

        printf("\nAccuracy <%d> == %f",kn,smthing);
        smthing=0;
        countot=0;

}
```

9.3 Sample code for main program

```
#include<stdio.h>
#include<string.h>
#include<stdlib.h>
#include <time.h>
#define maxf 20
void eucleid();
void manht();
void mink();
FILE *infile,*outfile,*out_train,*out_test,*result;
int test_count=0;
int train_count=0;
int i,j,k,kn,max,maxc,tempsl,maxclass=0;
int a=0;
int b=0;
int c=0;
int l;
int corrcount=0;
int countot=0;
int *sl,*wt,*wtlb,*count,*count_test;
float **arrtr;
float **arrtst,*arrtst1;
float **dist,**arr,***arry;
float temp,tot_dist=0;
float *tdist;
main()
{

        char *filename,*outw,*outtr,*outtst;
        filename=(char*)malloc(sizeof(char)*maxf);
        fflush(stdin);
        printf("Enter the infile name :: ");
        scanf("%s",filename);
        printf("\nInfile is  --> %s",filename);
        outw=(char*)malloc(sizeof(char)*maxf);
        outtr=(char*)malloc(sizeof(char)*maxf);
        outtst=(char*)malloc(sizeof(char)*maxf);
        printf("\n\nEnter the out file name :: ");
        scanf("%s",outw);
        printf("\nOutfile is  -->   %s",outw);
        printf("\n\nEnter training file name :: ");
        scanf("%s",outtr);
        printf("\nTraining file  --> %s",outtr);
        printf("\n\nEnter test file name :: ");
        scanf("%s",outtst);
        printf("\nTesting file --> %s",outtst);
        infile=fopen(filename,"r");
        outfile=fopen(outw,"w");
        out_train=fopen(outtr,"w");
        out_test=fopen(outtst,"w");
        result=fopen("rppr","w");
        srand((unsigned)time(NULL));
        char ch;
        int m,n,maxr;
        int ini,row=0;
        int column=1;
        int init;
        float per;
        ch=getc(infile);

        while(ch!='\n')

        {

                ch=getc(infile);
                if(ch==' ')
                {

                        column++;

                }

        }

        row=1;
//      fprintf(outfile,"\nRow read successfully\n");
        while(!feof(infile))
        {

                ch=getc(infile);
                if(ch=='\n')
                {

                        row++;

        }
        }

        for(i=0;i<maxr;i++)
        {
```

```
        for(j=0;j<maxc;j++)
        {

            fscanf(infile,"%f",&arr[i][j]);
            init=arr[i][0];
            if(init>maxclass)
            {

                maxclass=init;

            }
            else
            {

                maxclass=maxclass;
            }
        }

    }

    scanf("%d",&kn);

    printf("\nSelect the distance you want to calculate :: \n\t\t1 --> Eucleid\n\t\t2 -->
Manhattan\n\t\t3 --> Minkowski\n\t\t\t\t");

    scanf("%d",&i);
    switch(i)
    {

        case 1: if(sqrt(maxr)>train_count)
            {

                for(kn=1;kn<sqrt(maxr);kn++)
                {

                    eucleid();
                }

            }
            else
            {

                for(kn=1;kn<=train_count;kn++)
                {
```

```
                                        eucleid();
                                  }

}
                        break;
               case 2:  manht();
                        break;
               case 3:  mink();
                        break;
               default:  printf("\nPlease enter a valid choice");
       }
       fclose(infile);
       fclose(outfile);
       fclose(out_test);
       fclose(out_train);
       fclose(result);

}
```

9.4 HTML Code of GUI:

```
<html>
<head>
</head>
<body
bgcolor="lightgrey">
<font color=blue>
<h1><u>GRAPHICAL USER INTERFACE OF CLASSIFIER</u></h1>
</font>
<h2><u><font color=red>
PLEASE SELECT THE CLASS</U></FONT></H2>
<P>
<form action="data.php" method="get">
<select name="classname">
<option value="1">PRIVATE</option>
<option value="2">STATEGOV</option>

<option value="3">FEDERALGOV</option>
<option value="4">SELFEMPNOTINC</option>
<option value="5">UNKNOWN</option>
<option value="6">NEVERWORKED</option>
<option value="7">SELFEMPINC</option>
<option value="8">LOCALGOV</option>
</select>
<H2><U>CRITERIA FOR YOUR SEARCH:</U></H2>
<select name="choice">

<option value="1">SHOW ALL VALUES</OPTION>

<option value="2">ATTRIBUTE1</option>
<option value="3">ATTRIBUTE2</option>
<option value="4">ATTRIBUTE3</option>
<option value="5">ATTRIBUTE4</option>
<option value="6">ATTRIBUTE5</option>
<option value="7">ATTRIBUTE6</option>
<option value="8">ATTRIBUTE7</option>

<option value="9">ATTRIBUTE8</option>
<option value="10">ATTRIBUTE9</option>

<option value="11">ATTRIBUTE10</option>
<option value="12">ATTRIBUTE11</option>
<option value="13">ATTRIBUTE12</option>
<option value="14">ATTRIBUTE13</option>
<option value="15">ATTRIBUTE14</option>
<option value="16">ATTRIBUTE15</option>
```

```
</select>
<H3>INPUT SEARCH VALUE:
<INPUT TYPE="text" name="values" size="30">
<input value="SUBMIT" type="submit">
<input type="Reset">

</form>
</body>
</html>
```

CONCLUSION

Semi-supervised learning is of great interest in machine learning and data mining because it can use readily available unlabeled data to improve supervised learning tasks when the labeled data are scarce or expensive. Semi-supervised learning also shows potential as a quantitative tool to understand human category learning, where most of the input is self-evidently unlabeled.

The main goal of the project is to develop a more effective algorithm for semi-supervised learning, develop a tool for the same. It also aims to implement the same in any real life field like image processing, web mining, multimedia processing, medical, bio-informatics datasets, etc, if allowed by the time constraint.

BIBLIOGRAPHY

Books:

- Machine Learning, Tom Mitchell, McGraw Hill
- Han and Kamber: Data Mining---Concepts and Techniques

Research Papers

- Xiaojin Zhu and Andrew B. Goldberg. Introduction to Semi-Supervised Learning. Morgan & Claypool, 2009Eammon Keog, Time Series in SSL
- Xiaojin Zhu. Semi-Supervised Learning. Encyclopedia entry in Claude Sammut and Geoffrey Webb, editors, Encyclopedia of Machine Learning.
- Andrew Goldberg, Xiaojin Zhu, Aarti Singh, Zhiting Xu, and Robert Nowak. Multi-manifold semi-supervised learning. In Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS), 2009.
- Andrew B. Goldberg and Xiaojin Zhu. Keepin' it real: Semi-supervised learning with realistic tuning. In NAACL 2009 Workshop on Semi-supervised Learning for NLP, 2009.
- Xiaojin Zhu, Andrew B. Goldberg, and Tushar Khot. Some new directions in graph-based semisupervised learning (invited paper). In IEEE International Conference on Multimedia and Expo (ICME), Special Session on Semi-Supervised Learning for Multimedia Analysis, 2009.
- Aarti Singh, Robert Nowak, and Xiaojin Zhu. Unlabeled data: Now it helps, now it doesn't. In Advances in Neural Information Processing Systems (NIPS) 22, 2008.
- Xiaojin Zhu, Timothy Rogers, Ruichen Qian, and Chuck Kalish. Humans perform semi-supervised classification too. In Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07), 2007.
- Andrew Goldberg, Xiaojin Zhu, and Stephen Wright. Dissimilarity in graph-based semi-supervised classification. In Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS), 2007.
- Gregory Druck, Chris Pal, Xiaojin Zhu, and Andrew McCallum. Semi-supervised classification with hybrid generative/discriminative methods. In The Thirteenth

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2007.

- Andrew Goldberg and Xiaojin Zhu. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing, New York, NY, 2006.
- Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Department of Computer Sciences, University of Wisconsin, Madison, 2005.
- Maria-Florina Balcan, Avrim Blum, Patrick Pakyan Choi, John Lafferty, Brian Pantano, Mugizi Robert Rwebangira, and Xiaojin Zhu. Person identification in webcam images: An application of semi-supervised learning. In ICML 2005 Workshop on Learning with Partially Classified Training Data, 2005.
- Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Semi-supervised learning: From Gaussian fields to Gaussian processes. Technical Report CMU-CS-03-175, Carnegie Mellon University, 2003.
- Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.