

CMPE 255-Team Project Proposal

Project Title : Income Classification

Team Information :

Team member Name	SJSU ID	Email ID
Bhavana Bangalore Sathyaprakash	014597245	bhavana.bangaloresathyaprakash@sjsu.edu
Raaga Pranitha Kolla	011824410	raagapranitha.kolla@sjsu.edu
Namratha Bhat	014597258	namratha.bhat@sjsu.edu

Dataset :

The dataset we are using for this project is [Census-Income \(KDD\) Data Set](#). This data is collected from UCI machine learning repository website. The dataset contains 299285 records and 40 attributes. This is a multivariate data consisting of integer and categorical values. For testing, the dataset we use contains 99762 records and 40 attributes.

Filename	File Description
census-income.data	Used for training (199523 records)
census-income.test	Used for testing (99762 records)
census-income.names	Contains description about each attribute in the dataset.

Project Description :

This project focuses on implementing the following stages of data pipeline such as data preprocessing, exploratory data analysis, data cleaning, implementing and evaluating multiple classification algorithms such as Naive Bayes, SVM, logistic regression, etc . We plan to implement a minimum of three types of classifications(one per person), if time permits, we will work on other algorithms. At the end, all the results would be compared based on the prediction performance.

Tentative project plan :

Task	Responsibility	Execution
Exploratory Data Analysis	All	We plan to have weekly meetings to discuss our

Data Preprocessing	All	progress and setbacks.						
Modelling	<table><tr><td>Bhavana Bangalore Sathyaprakash</td><td>Naive Bayes (tentative)</td></tr><tr><td>Raaga Pranitha</td><td>SVM(tentative)</td></tr><tr><td>Namratha Bhat</td><td>Logistic Regression(tentative)</td></tr></table>		Bhavana Bangalore Sathyaprakash	Naive Bayes (tentative)	Raaga Pranitha	SVM(tentative)	Namratha Bhat	Logistic Regression(tentative)
	Bhavana Bangalore Sathyaprakash		Naive Bayes (tentative)					
	Raaga Pranitha		SVM(tentative)					
Namratha Bhat	Logistic Regression(tentative)							
Report	All							

References :

1. <https://archive.ics.uci.edu/ml/datasets/Census-Income+%28KDD%29>
2. <https://scikit-learn.org/stable/modules/svm.html>
3. <https://ieeexplore.ieee.org/abstract/document/8748528>
4. <https://www.imperva.com/learn/data-security/data-classification/>