

EL-GY 6123
Introduction to Machine Learning
PROJECT PROPOSAL

Project Title: “Amazon-Employee Access Prediction.”

Students in the Group:

Akhil Kumar Dundigalla, akd389, N14286972

Sravan Reddy Chintareddy, src572, N15762451

Sridhar Gadicherla, sg5764, N15766107

Proposal :

When any employee at any company starts to work, they want to obtain the complete computer access necessary to fulfill their role. This access may allow an employee to manipulate resources according to his needs. But often many employees figure out that their access is restricted to certain applications. A supervisor then manually gives the required access to the employee in order to overcome access obstacles which takes a lot of time. As employees move throughout a company, this manual procedure wastes a nontrivial amount of time and money. There is enough amount of data regarding an employee's role within an organization and the resources to which they have access. Given this data, models can be built that automatically determine whether access privileges can be given to an employee or not.

Our objective is to build a classification model, learned using historical data provided by amazon and available from Kaggle competitions, that will determine whether an employee should be given access or not. Since it is a classification problem we will first try to use Linear classifiers (SVM, Logistic Regression) and measure the performance using Accuracy or ROC-AUC. If the performance is not up to the mark then we will try to use RandomForest classifier. We will use the best model we can come up with and use it to predict what is required and measure its performance using ROC-AUC.

Data: The data consists of real historical data collected by Amazon from 2010 & 2011 and is available at <https://www.kaggle.com/c/amazon-employee-access-challenge/data>