**HR Retention Prediction**

**Project Report**

Submitted by

**Arpit Khurana (15BCE0353)**

**Tanmay Jha (15BCE0618)**

**Faheem Ali (15BCE0740)**

In fulfillment for the project of

**Social and Information Networks**

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING





**School of Computer Science and Engineering**

**DECLARATION**

We hereby declare that the project entitled HR Retention Prediction submitted by us to the School of Computer Science and Engineering, VIT University, Vellore-14 in fulfillment of the requirements for the award of the project of social and information networks in Computer Science and Engineering is a record of bonafide work carried out by us under the supervision of Annapurna Jonnalagadda, Assistant Professor. We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other project of this institute or of any other institute or university.

**Arpit Khurana (15BCE0353)**

**Tanmay Jha (15BCE0618)**

**Faheem Ali (15BCE0740)**



**School of Computer Science and Engineering**

**CERTIFICATE**

The project report entitled HR Retention Prediction is prepared and submitted by Arpit Khurana (Register No: 15BCE0353), Tanmay Jha (15BCE0618) and Faheem Ali (15BCE0740). It has been found satisfactory in terms of scope, quality and presentation as partial fulfillment of the requirements for the award of the project of Social and Information Networks in Computer Science and Engineering in VIT University, India.

Annapurna Jonnalagadda

Assistant Professor

**Guide**

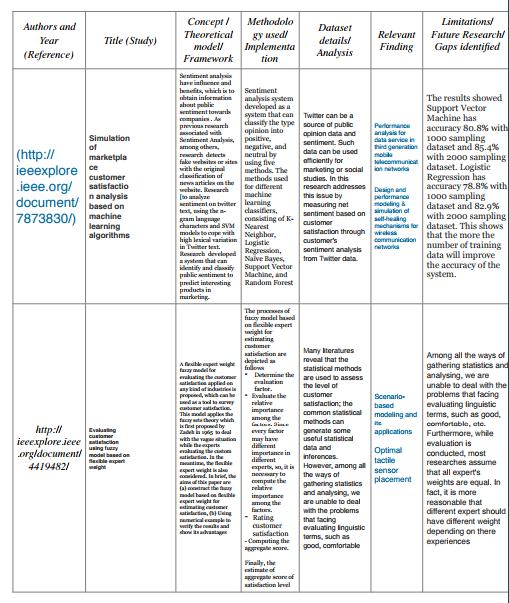
**Problem​ ​statement**

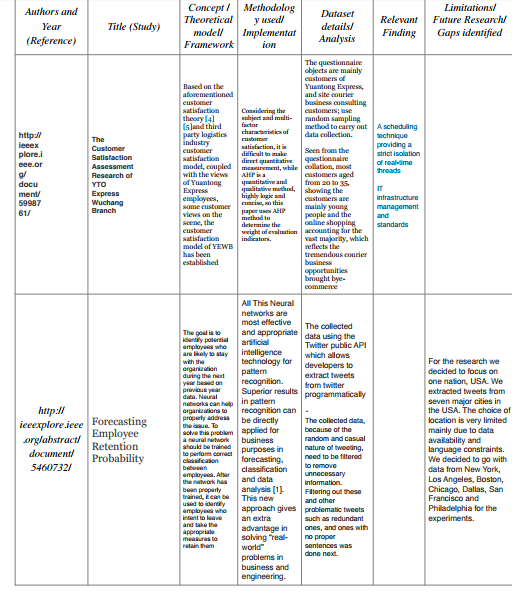
Employee attrition is very important. Companies spend lots of resources, money and valuable time to train their employees to get the best results and hence if the employee leaves the company due to some reason it is a great loss of human resource for the company. Therefore, understanding why employees leave is very important for businesses to maximize efficiency and reduce training cost which translates into greater profit.

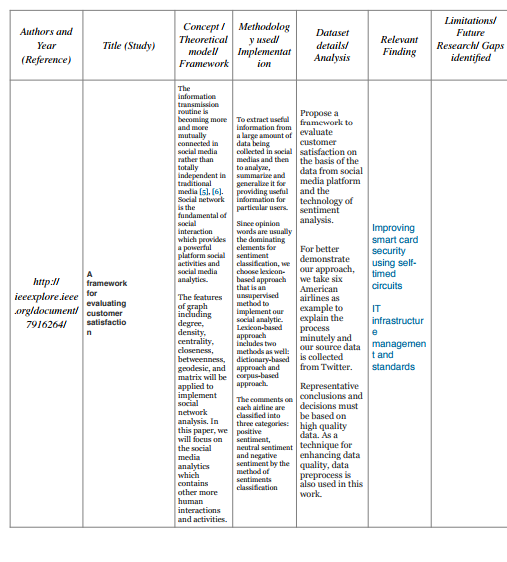
**Problem Significance and Application**

Organizations spend heaps of assets, cash and profitable time to prepare their representatives to get the best outcomes and thus if the worker leaves the organization because of some reason it is an extraordinary loss of human asset for the organization. In this manner, understanding why representatives leave is vital for organizations to expand proficiency and lessen preparing cost which converts into more prominent benefit. This additionally helps in making a friendlier and vivacious work space for representatives to work in which brings about better efficiency and imparts a feeling of gratefulness for what work they are doing. Through this examination, we might want to think of the most ideal orders which could be utilized to anticipate if a specific representative will take off. From our analysis we will find what type of people leave their job. And what are the values of different parameters for these employees like they have high salary or low. With the help of our data companies can take steps to prevent their employees from leaving the company.

**Supporting Document**







**Methodology:-**

**a. Algorithms and Approach:**

In our given dataset, we have data about the relations of every employee with each other. So, we first plot a graph which helped us to understand the different communities or group that has been built within the organizations. In our graph, the circles would represent people, and an edge between two vertices signifies that those two individuals are related to each other. This graph helped us understand if any employee feels left alone or if some employee is working with those people who are not within his domain of work. Then we also have the data of the employee about their satisfaction level, Last evaluation, Time since last performance evaluation (in Years), Number of projects completed while at work, Average monthly hours at workplace, Number of years spent in the company, Whether the employee had a workplace accident, Whether the employee left the workplace or not, Whether the employee was promoted in the last five years, Department in which they work for. We have used pandas to load the data in tabular form. Then we obtained the number of employees who have left the company. We plotted this in a pie chart. Then we found the satisfaction levels of employee and plot it on bar graph. We also, plotted the number of work accidents by users. We plotted the number of employees who were going to be promoted and who decided to leave their department. From, these graphs and plots we concluded what type of people left their job. And what are the values of different parameters for these employees like they have high salary or low. Finally, from this data, companies can take steps to prevent their employees from leaving the company.

**b. Dataset:**

Link of dataset: (<https://www.kaggle.com/ludobenistant/hr-analytics/discussion/26565>)

**c. Tools and Frameworks:**

**Language:** Python

**I) Beautiful Soup:** Beautiful Soup provides a few simple methods and Python idioms for navigating, searching, and modifying a parse tree using Python parsers like lamp and html5lib. It automatically converts incoming documents to Unicode and outgoing documents to UTF-8.

**ii) Text Blob:** Text Blob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

**iii) Pandas:** pandas is a Python package providing fast, flexible, and expressive data Page 7 of 16 structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

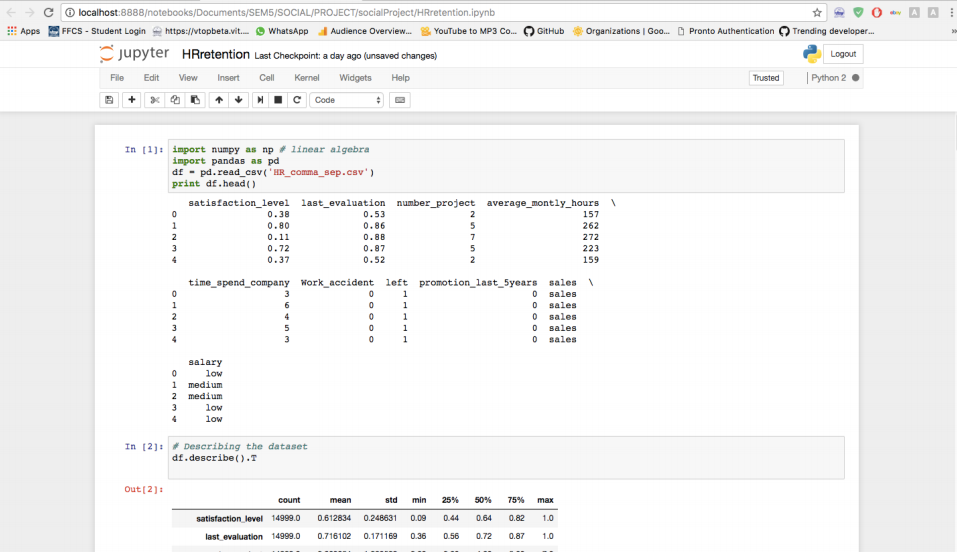
**iv) NumPy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

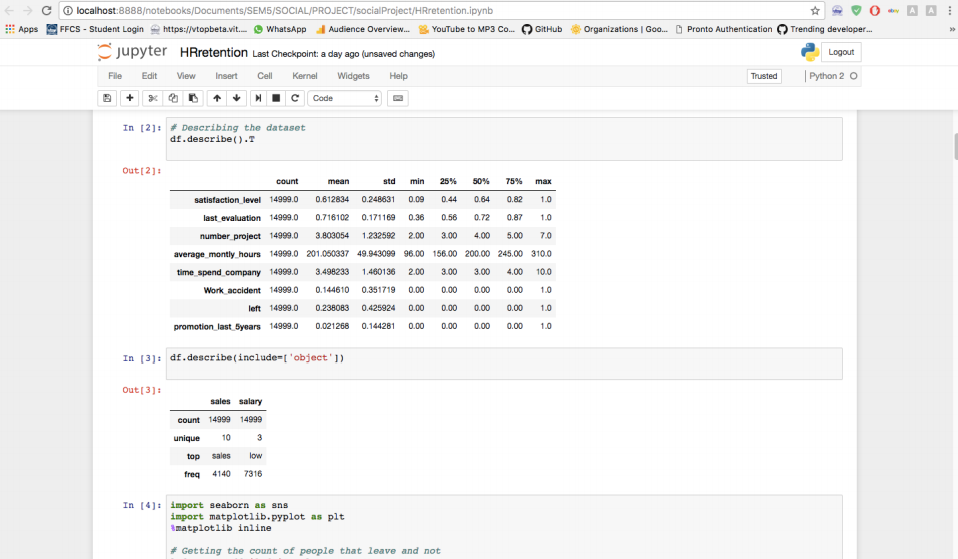
**v) NetworkX: -** Parallel Betweenness: To determine who the most important individual in the network is. By using the concept of Betweenness Centrality. Community Detection to identify different communities/clusters in the social network

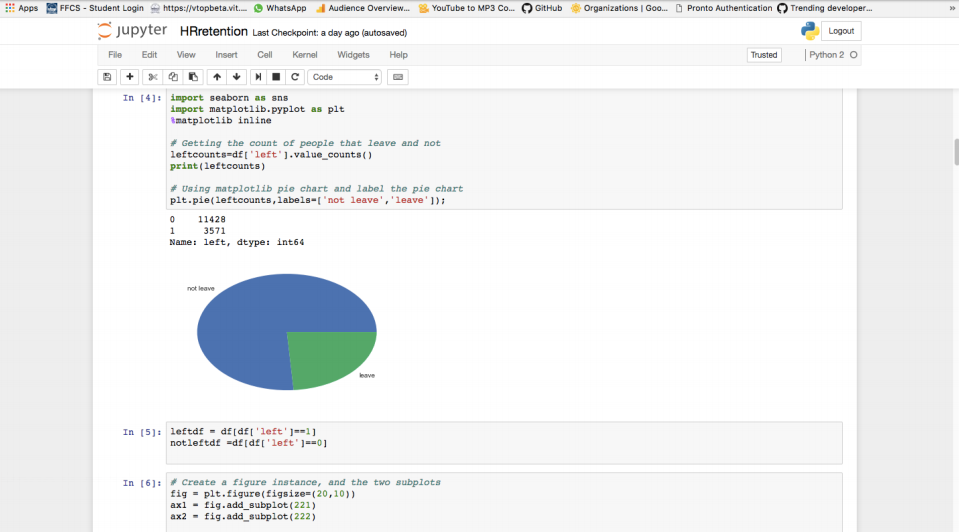
**vi) Anaconda: -** Anaconda is the world’s most popular Python data science platform. Anaconda, Inc. continues to lead open source projects like Anaconda, NumPy and SciPy that form the foundation of modern data science. Anaconda’s flagship product, Anaconda Enterprise, allows organizations to secure, govern, scale and extend Anaconda to deliver actionable insights that drive businesses and industries forward.

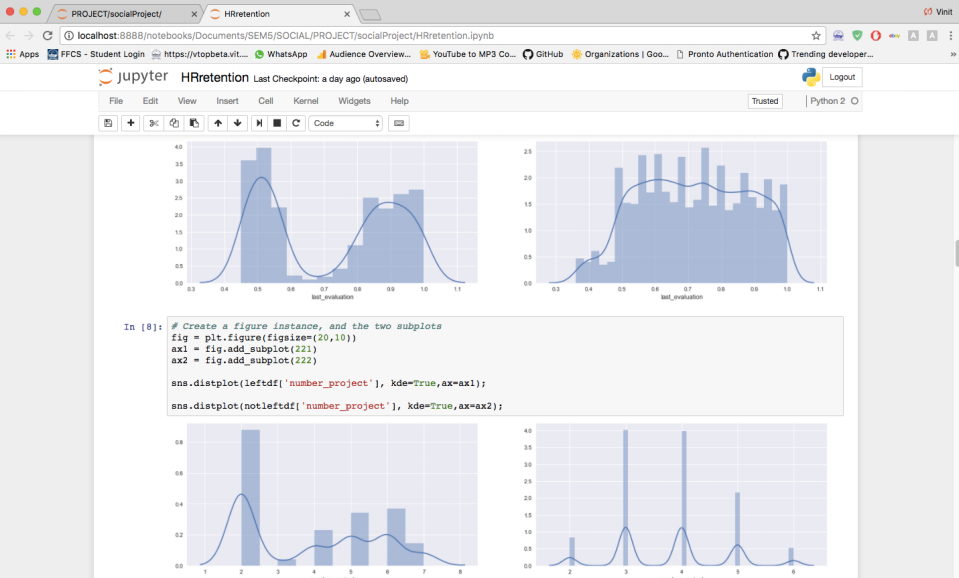
**vii) Jupyter :-** Notebook documents (or “notebooks”, all lower case) are documents produced by the Jupyter Notebook App, which contain both computer code (e.g. python) and rich text elements (paragraph, equations, figures, links, etc...). Notebook documents are both human-readable documents containing the analysis description and the results (figures, tables, etc...) as well as executable documents which can be run to perform data analysis.

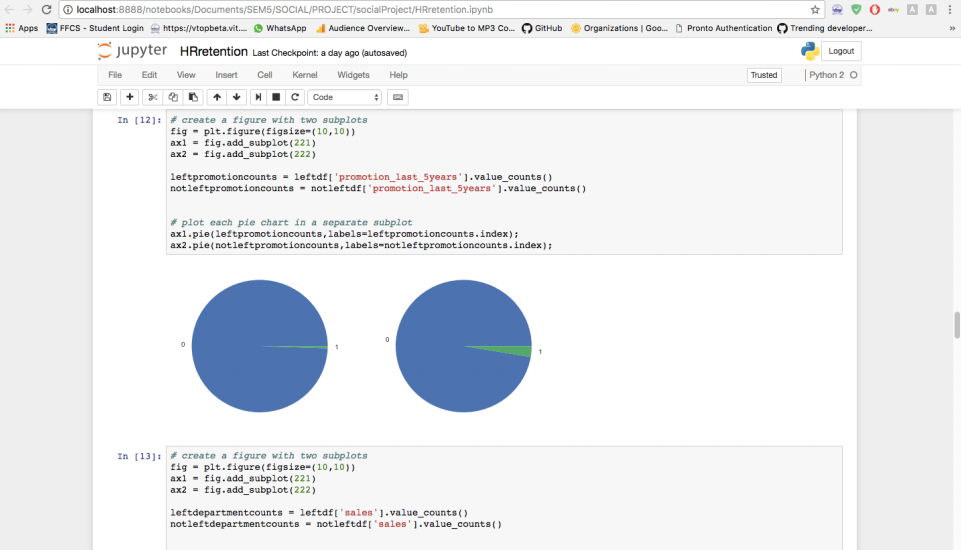
Output and demo

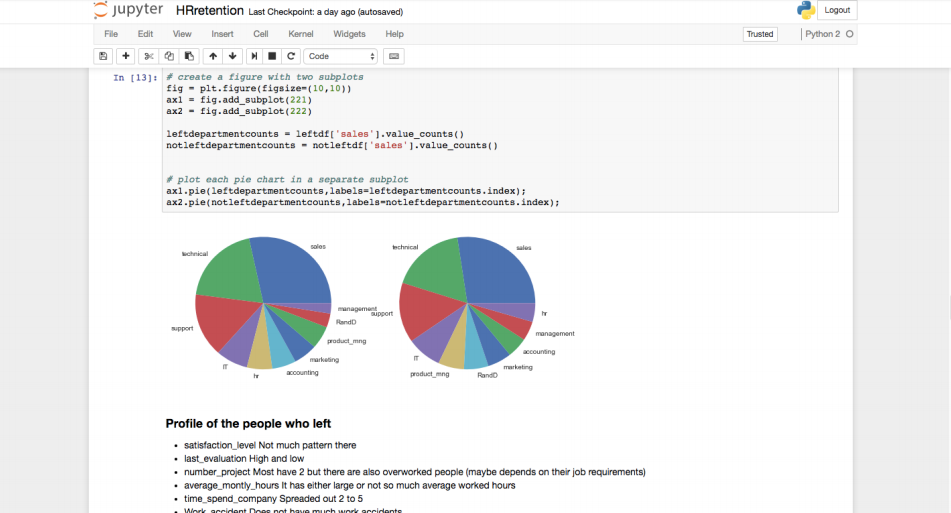


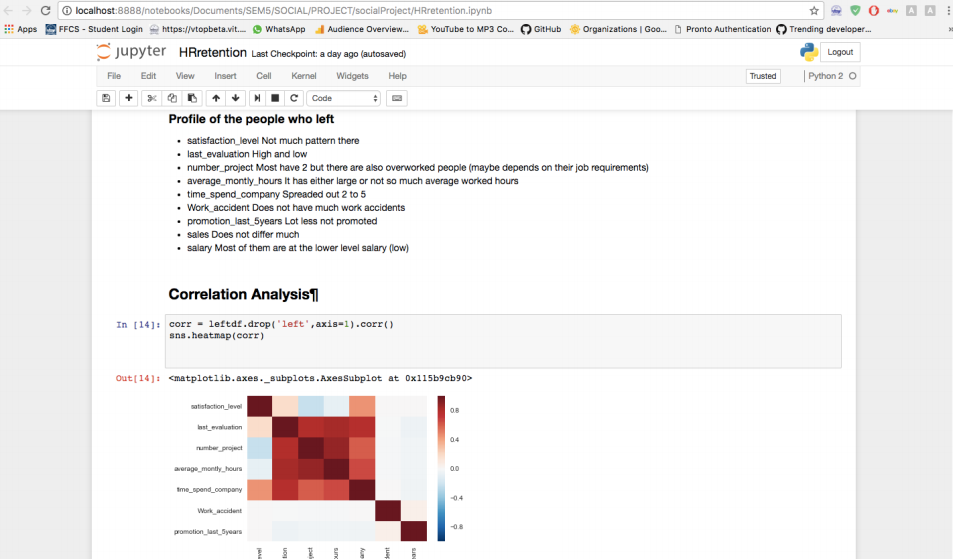


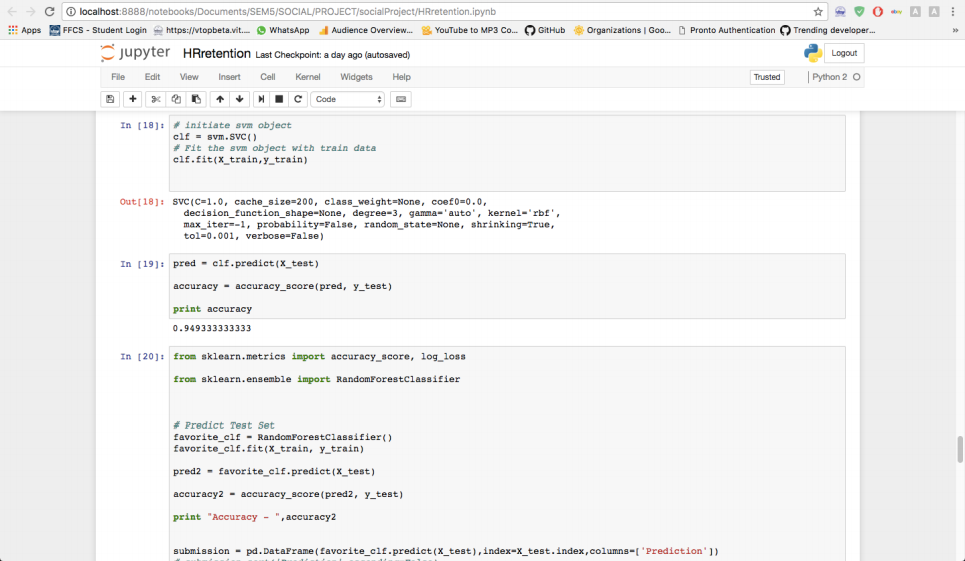


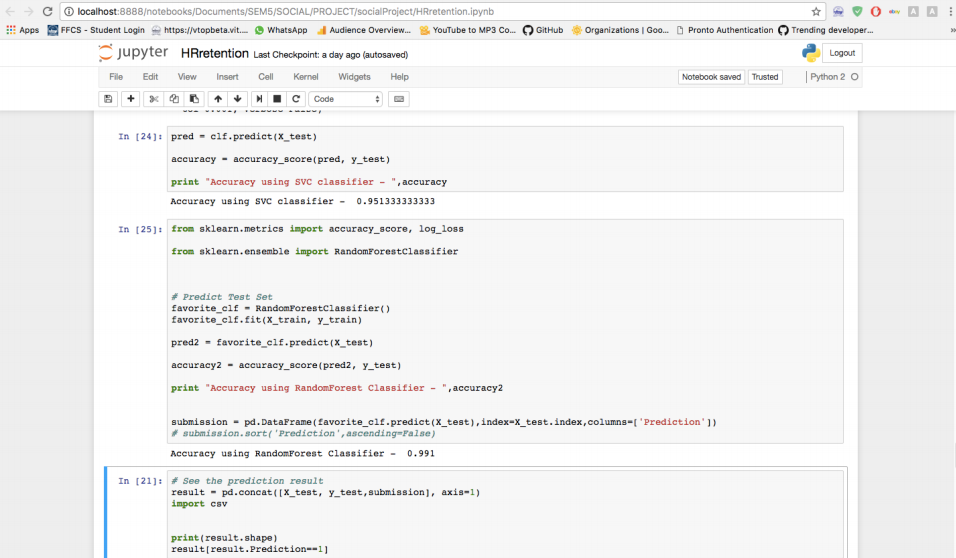


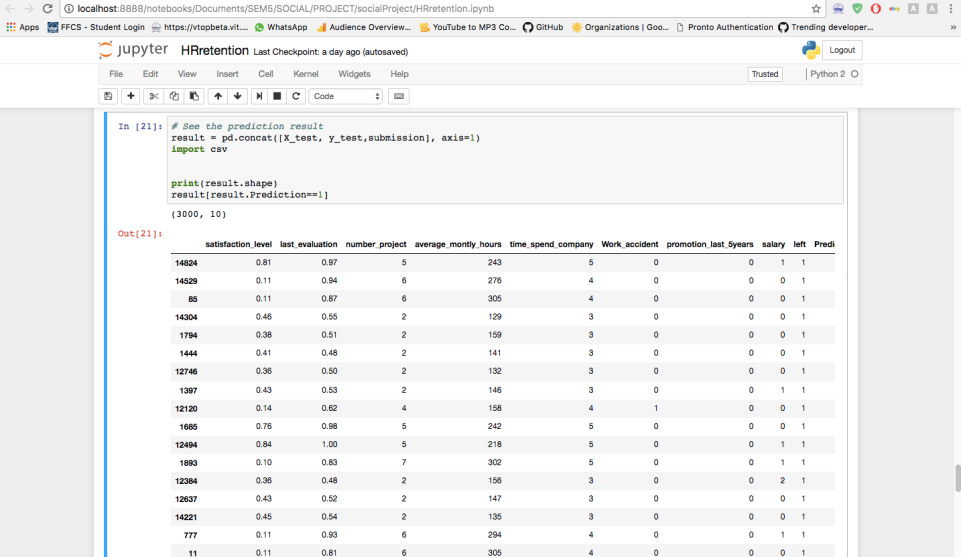


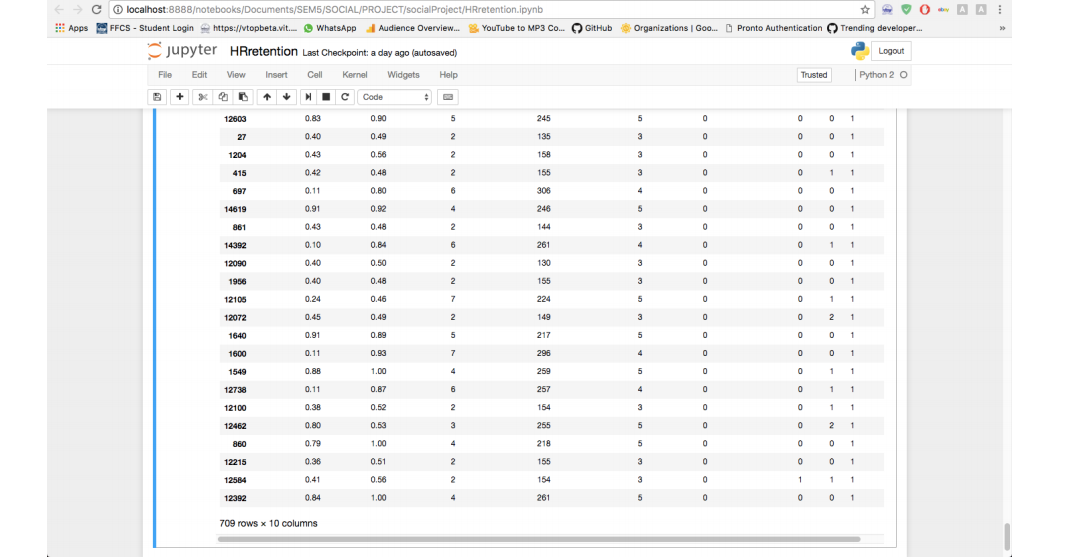




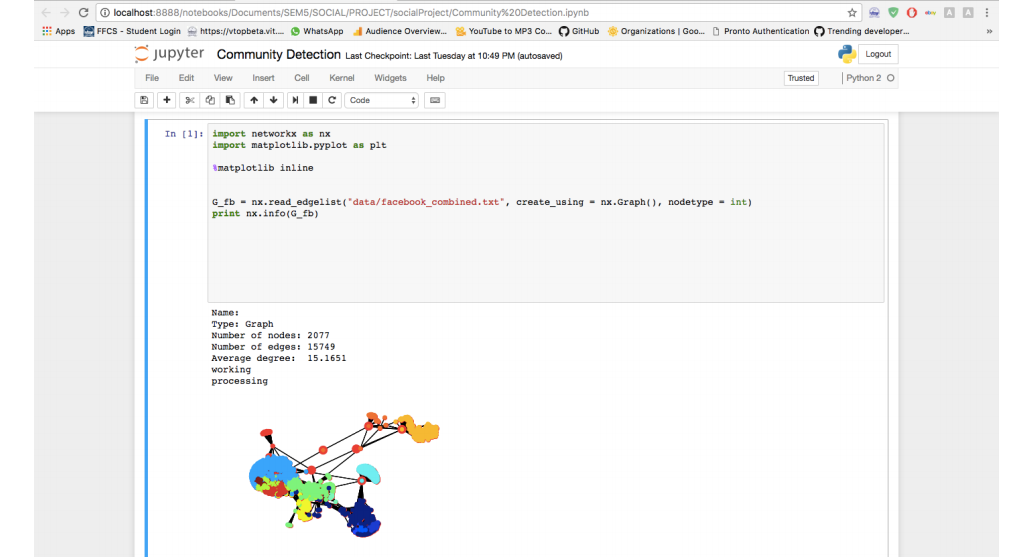


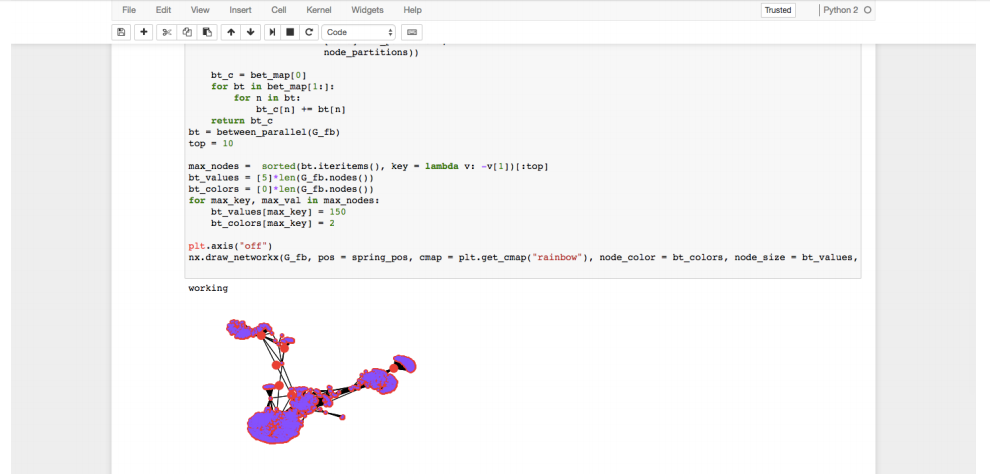


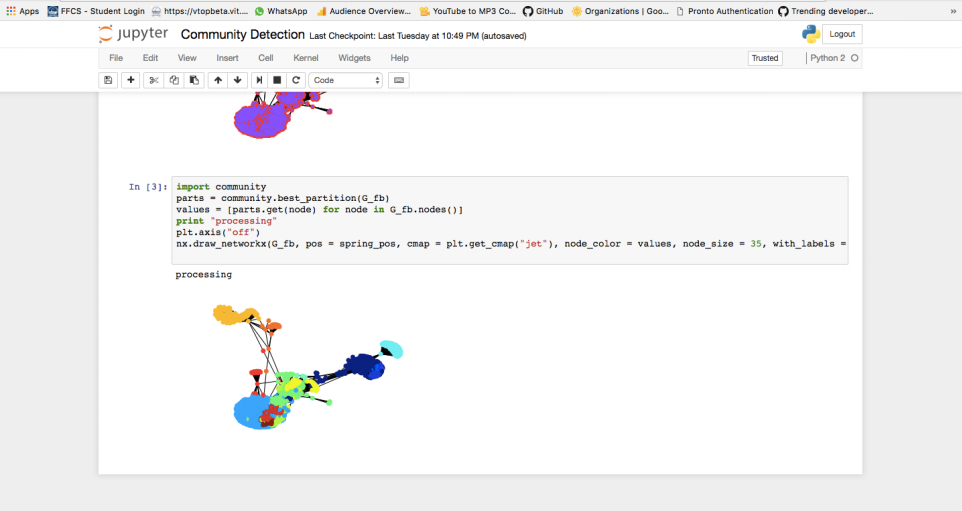




Community Detection and Betweenness Centrality :







**Results and discussion**

• Analysis of given Dataset

• Draw graphs for feature extraction

• Train model on the given dataset using SVC and Random forest Classifiers

• Predict whether the employee is going to leave job or not, based on our trained model.

• Identified what may be the consequences, he is facing, which will be useful for employee retention.

• Perform and calculate community detection and betweenness of employee(who is thinking of leaving the job), so that we can talk to them and help the employee as well as company both.

**Effect on the original problem**

Organizations spend heaps of assets, cash and profitable time to prepare their representatives to get the best outcomes and thus if the worker leaves the organization because of some reason it is an extraordinary loss of human asset for the organization. In this manner, understanding why representatives leave is vital for organizations to expand proficiency and lessen preparing cost which converts into more prominent benefit. This additionally helps in making a friendlier and vivacious work space for representatives to work in which brings about better efficiency and imparts a feeling of gratefulness for what work they are doing. From our analysis we found what type of people leave their job. And what are the values of different parameters for these employees like they have high salary or low. With the help of data companies can take steps to prevent their employees from leaving the company using algorithms in machine learning. After that, we performed community detection algorithm and calculate betweenness centrality of those persons, who are thinking of leaving their job, so that we can contact to people, who are surrounded by those, who are leaving their jobs.

**References :**

i) Serrat, Olivier. "Social network analysis." Knowledge solutions . Springer Singapore, 2017. 39-43.

ii) Marcus, S., Melanie Moy, and Thayne Coffman. "Social network analysis." Mining graph data (2007): 443-467.

iii) McGloin, Jean Marie, and David S. Kirk. "Social network analysis." Handbook of quantitative criminology . Springer New York, 2010. 209-224.