Shri Vishnu Engineering College for Women (AUTONOMOUS):: Bhimavaram

Department of Computer Science & Engineering

Academic Year: 2022-23

IV / II B. Tech - Project Work

ABSTRACT

Name of the Class / Section	IV-II B. Tech CSE-C		
Batch Number	C14		
Project Domain Area	MACHINE LEARNING		
Project Title	EMPLOYEE CHURN PREDICTION AND ANALYSIS USING MACHINE LEARNING ALGORITHMS		
Guide Name	Mr. G. RAMESH BABU		
Students Registered	**Registered Number**	**Student Name**	**Student Signature**
	19B01A05F6	PUPPALA VENKATA LAHARI	
	19B01A05G8	TALAPULA ASWITHA	
	19B01A05H9	VEGESNA NIKITHA DEVI	
	19B01A05I8	YADAVALLI SATYA VENKATA ANITHA	

Abstract of the Project (In 200 words)

A person working for an organization is the vital resource which is known as an employee. If one of them leaves suddenly, this could affect and cost massive amount to respective company. And recruitment would consume not only time and money but also the newly joined person needs some time for making particular business cost-effective. While it has become apparent that employee churn prediction responds differently to salary, depending on their location, lifestyle, and environment, the linked knowledge and understanding remain fragmented. In this paper, we aim to design expert prediction system to deal with problems associated with lack of knowledge of employee behavior, to aware organizations about the importance of employee, to prevent unnecessary employee churn, and to improve growth of both separately.

Employee churn is a major concern for large companies as there is loss of valuable resource and extra resources like time and money has to be spent on recruiting new resources which directly affects their revenues. Due to this, companies are seeking means to predict factors that influence employee churn. The main goal of our work is to develop a churn prediction model that assists employers in predicting which employees are most likely to be subject to churn.

The model developed in this work uses machine learning techniques to identify the underlying patterns so that they can be maneuvered to reduce churn. The model experimented with seven algorithms: Logistic regression, Random Forest, Bagging Classifier, KNN, Gradient Boost, XgBoost, and SVM.

Existing System – Drawbacks

1.Present churn predictors mainly focus on only customers but not on employees.
2.Most of the Existing systems use mainly logistic regression and KNN algorithms only.
3.current system uses around 20 parameters for prediction.
4.Low Accuracy

Proposed System – Features
List of objectives/features that are planned to implement.

- Testing the data set with different classification and clustering algorithms.
- Testing different data sets with different classification algorithms.
- Comparing the results of different classification algorithms.
- Predicting the result using best accurate model.

ADVANTAGES:

1. present churn predictors mainly focus on only customers but not on employees so our model focus on employee churn.
2.Most of the Existing systems use mainly logistic regression and KNN algorithms only as an algorithm whereas our model uses 7 algorithms
3.current system uses around 20 parameters for prediction and our model uses around 35 parameters

(i)Functional Requirements

- User can provide a csv file containing information of required attributes
- The trained model will predict the employee retainment in a company over a certain period of time.
- After the prediction the user can search for the EMPLOYEEID whether he stays or leaves
- User can also download csv files of:
  1.Entire report
  2.Leaving employees
  3.staying Employees
- User can also learn information regarding the algorithms we use for our model
- An entire page dedicated to find the relation between the employees' attrition and other features

(ii) Non - Functional Requirements

- Performance
- Usability
- Reliability
- Accuracy
- Maintainability

(iii) Software Requirements

- Python version above 3
- Jupyter notebook
- Vs code
- Html, CSS
- Libraries required:
  1. Pandas
  2. TensorFlow
  3. Scikit-learn
  4. matplotlib
  5. seaborn
  6. Flask

(iv) Hardware Requirements

- Operating system: windows 10 or above
- Ram: 4GB
- Hard disk: 32 GB
- Architecture:x86 64-bit CPU (Intel / AMD architecture) for python
- Memory and disk space required per user: 1GB RAM + 1GB of disk +. 5 CPU core.
- Server overhead: 2-4GB or 10% system overhead (whatever is larger). 5 CPU cores.
- Port requirements: Port 8000 plus 5 unique, random ports per notebook.

Literature Survey	**A survey of employee and customer churn prediction methodologies Year – 2020** Various classification models in data mining can be used for churn prediction. This research focuses on surveying the various such prediction methods like decision trees, logistic regression, K-Nearest Neighbor (KNN), feature selection and comparing the performance of these nine methods in terms of their average accuracy. **Churn Prediction using Supervised Machine Learning algorithms -Impact of Oversampling Year – 2020** Customer data can be used to predict the customer behavior using machine learning algorithms and the telecommunication industries receive a lot of customer data every day. We perform supervised machine learning algorithms to predict customer churn along with taking into consideration the challenges that are faced during the development of the prediction models like), K-Nearest Neighbor (KNN), Random-forest, Logistic Regression and XgBoost.
Modules	**Expected date of completion**
Web Scrapping and Installations	23-11-22
Data Cleaning, feature Engineering and Exploratory data analysis	20-12-22
Model Building	20-01-23
Model deployment	05-02-23
Testing the project	05-02-23
Final Project Report	13-02-23

Signature of Internal Project Guide	**Signature of C Sec :: Project – Coordinator**	**Signature of HoD- CSE**