Project :

Employee churning identification

BY

PRANJAL SHUKLA

Year : 2020

- **Introduction:**

The project “Employee churning identification” ANN algorithm which will help determine whether a employee should be churned or not.

This project will help the CEO’s or higher authority to decide whether an employee should be churned or not.

This project model can be used in private as well as public sector for efficient working of organization.

- **Understand and define the problem :**

I. Problem Statement :-

Using ANN algorithm decide whether an employee should be churned or not.

II. Objectives :-

a. To identify whether an employee should be churned or not
b. To find accuracy of model

III. Scope :-

a. This model can be used by authorities of organization.
b. This model can be used in the both private and public sector for efficient operation.

IV. Data Sources:

The various data sources used are:

a. CSV data
b. Excel data
c. MySql data

V. Tools and Techniques :

The various tools and techniques used are:

a. Pandas library
b. Numpy library
c. Matplotlib and seaborn library for data visualization
d. Sci -kitlearn library for deep learning algorithm.

- **Dataset Preparation and Preprocessing**

In this stage of project implementation, focus is put on data collection, data selection, data preprocessing, and data transformation.

I. Data collection :-

The data used in this project is in the form of csv file consisting of rows and columns in the form of table.

II. Data Visualization :-

In this project we are using seaborn and matplotlib library of python for data visualization.

The graph plotted is a countplot to compare number of employees churned and not churned.

III. Labeling

As supervised learning is implemented in the model we have a target column in our dataset named “Exited” which consist of values “0” meaning the employee will not be churned and “1” for the employee to be churned

IV. Data Selection:-

The dataset used in the project contains insignificant columns which need to be dropped like : CustomerID , RowNumber,Surname,Geography,Balance

V. Data Preprocessing

The purpose of preprocessing is to convert raw data into a form that is useful in training and

testing the DL model. The structured and clean data produces more precise results. In short,

good quality data when fed to the DL model, it produces better results.

The Preprocessing technique includes data formatting, cleaning, and sampling techniques.

a. Data Formatting : -

   We have used the Standard Scaler function to normalize our dataset

b. Data cleaning:

   As our dataset did not contain null values for numerical columns and it also did not contain null values after converting categorical columns to numerical, our dataset is clean

c. Data anonymization: -

   The dataset used in project do not contain any sensitive information.

d. Data Transformation:-

We have converted categorical columns to numerical columns using the LabelEncoder function .

Scaling: The dataset columns are normalized using StandardScaler function

Feature Extraction: Some of the existing features are combined to create new features which are useful for ML modeling.

VI. Dataset Splitting:

The given dataset is split into three parts: training, testing, and validation sets. The ration of training and testing sets is typically 80 to 20 percent. The 20 percent of the training set is further split as a validation set.

- ## Model Training:-

In this stage, the training data is fed to the Deep learning algorithm to build and train.

The different algorithms used to train the models are:

1. Artificial Neural Network

- ## Model Testing and Evaluation

The goal of this step is to develop the simplest, reliable and efficient model.

We have used the above mentioned algorithms to train the model and also calculated accuracy on training as well as testing data.

From the results, We can say that the ANN algorithm gives an accuracy of 0.815 on testing data and accuracy of 0.8615 on training data.

- ## Improving Predictions with Ensemble Methods

As the Decision Tree algorithm provides the best accuracy for our model we are using the following technique to improve predictions:

1. Bagging: In this case, the models of the same type are combined in sequential manner. The training dataset is split into subsets. Then the models are trained on each of these subsets.

- **Model Deployment**

We are using the Web Service based Deployment in which the prediction is done continuously. Mostly the private or public cloud is used for deployment.

- **Conclusion and Further Development:**

This project explains the stages of machine learning used to make the model which mainly are:

1. Data Collection
2. Data Preprocessing
3. Data visualization
4. Splitting the dataset
5. Training the model
6. Testing the model
7. Deployment of model

I would like to make changes in model which will help to know the criteria on which an employee should be churned ,which will provide an understanding to employee to know the areas where they lack.

.