Idea/Approach Details

Technology Bucket: Finance

Category: Software

Company Name/ Ministry Name: FIS Solutions(India)

Problem Code : AK2

Team Leader Name : Ashish Bhawnani

College Code : 1-3763789271

An income tax fraud detection idea using AI & ML

The task given in the challenge was to detect tax fraud using machine learning. We built a robust prototype which is live on web at
http://taxdetection.herokuapp.com/ .
(Username: fis Password: fis)
Inspection authorities have details of people like investments, bank account deposits generally found using PAN Number. These details and previous tax fraudsters data (sample data) is taken and a machine learning model is trained using Neural network algorithm. A default test file of people is built in prototype which can be used to check our prototype. When you click on predict in default file section in our prototype, you will see Graphs for each Pan Number showing Probability to Do Tax Fraud and Tax Fraud Contributing Reasons.

The backend of software prototype is built in Python with Scikit Learn and Keras Libraries. The backend is embedded into Flask framework for allowing interaction of python with frontend. The frontend is created in HTML, CSS and Javascript for displaying data in rich format and Graphs. Flask directory rules are followed so Flask can render the html files. Procfile and Requirements file is created and Git and Heroku CLI is used to deploy the prototype live on web. We created Flask framework enabled Python environment on Heroku. The algorithm which does prediction is based on Convolution Neural Network which has most amazing accuracy.

Idea / Approach details

This prototype has lot of practical use when trained on large original data. Using PAN Number of a person, authorities can extract his financial information. This information can be fetched to prototype. It will predict the probability of that person to do tax fraud depending on his financial information provided. The factors/ attributes used to predict are Bank Deposits, Income Revealed, Investments in Mutual Funds, Investments in Gold ETFs, CIBIL Score, Area of residence, Number of cars, Number of Houses. One important thing is to train the prototype model on real world data because on sample data it just predicts on basis of few sample entries. Also prototype predicts the reasons and factors contributing to our algorithm's calculation in that person's probability.

Our prototype can be accessed anywhere live on web because it is already deployed. The deployment was possible due to proper technology languages combination. This shows our prototype is highly scalable. Model is developed in Convolution Neural Network which has highest accuracy as proved by experts. Pre processing , Scaling of Data is also done before training the model and once trained on big data, the prototype will become real working software. Python backend with Flask framework makes it scalable, database connectable and deployable on any platform.

SCREENSHOTS: