RAVI KANT RAI

DATA SCIENTIST

CONTACT

rkrai2805@gmail.com
8860189093
Gurugram, Haryana

EDUCATION

B.Tech
Computer Science
Dr. A.P.J Abdul Kalam Technical University
2011 - 2015
Lucknow, Uttar Pradesh

GENERAL SKILLS

Machine Learning
Statistical Analytics
Data Preparation
Data Analysis
Data Visualization

Languages

Python, R Programming, SQL

Packages

Pandas, Numpy, NLTK, Scikit-learn, SpaCy, Pytesseract, geopandas etc

PROFILE

A Data Scientist with total **5+ years** of experience in IT domain, experience in executing data-driven solutions to increase efficiency, accuracy, and utility of internal data processing. Experienced at collecting, analyzing, and interpreting large datasets, developing new forecasting models, and performing data management tasks.

WORK EXPERIENCE

Data Scientist

Goods and Service Tax Network

July 2020 - current / Aerocity, Mahipalpur, Delhi

- Developed a logic to catch suspicious vehicle movements by using Eway bill part B dataset.
- Built machine learning model to predict the Risk Score for RFD01(Inverted Tax Structure under the GST) Refund Applications.
- Developed a model using softmax regression to suggest relevant HSN codes as per the goods/service name or description entered by the user.
- Developed a ML model to catch Fraud Taxpayers by using a supervised machine learning model Logistic Regression.

Data Scientist

Pernod Ricard

Aug 2019 - July 2020 / DLF Cybercity, Gurugram

- Developed and implemented new forecasting models which increased company productivity and efficiency.
- Build a document classification model, to labialize .pdf files into their categories according to their contents by using NLP and Unsupervised Machine Learning

Data Analyst

Citibank

July 2018 - Aug 2019 / Udyog Vihar Phase 3,Gurugram

- Consulted and worked with development team to determine, execute, and deliver relevant solutions.
- Provided information, feedback, and guidance to clients to support technology-related decision making
- Analyzed old information architectures and contributed to the design and development of the new one

L2 Engineer

Wipro Infotech

Aug 2016 - Jul 2018 / Gurugram

- Providing solutions and problem resolution to customers, with a strong sense of commitment and drive towards customer service.
- Collaborating with all levels of senior management at multiple organizations to meet organizational IT goals and expectations
- Developing or refining organizational performance measurement or security metrics.

PROJECTS

Refund Risk Score

- Designed a classification model to predict the risk score of RFD-01 Refund Applications(Inverted Tax Structure under the GST).
- To build this model used historical data of RFD01 Refund applications, and GSTIN network properties, with Income tax returns data. Prepared the dataset by cleaning and scaling
- Applied logistic regression to predict the Risk Score of the Refund Applications.

Search HSN code by Product name or Description

- Developed a ML based model to suggest relevant HSN codes as per the goods/service name or description entered by the user
- prepared a data dictionary for the model using historical data, cleaned up the raw data to create a data dictionary, and then converted this data from textual format to tabular format by using FastText.

Document Classification

- Using the document classification technique, we need to label .pdf files into categories according to their contents
- Extracted content from editable and scanned pdf files by using tesseract ocr, and then applied Data Cleaning techniques (Lemmatization, convert to lowercase, removal of stopwords and special characters etc.), and after that used feature Engineering techniques (TF-IDF,Word2Vec) to convert data from textual format to tabular format.
- Applied unsupervised machine learning algorithms on tabular data to create the desired clusters of documents, so that document retrieval can be accelerated.

Fraud Taxpayer Detector

- Build ML model or calculating the Probability to default by GST Tax payers.
- Used data of the GST-R1, GST-R2A and GST- R3B, PAN, Aadhar to build a ML model, to catch those taxpayers. Who may be defaulter in future.
- Build a scoring model by using a supervised machine learning model Logistic Regression