Arijit Hazra

Data Scientist

Mobile: +91 9689712560

Email: arijit.hazra23@gmail.com

Github: github.com/arijithazra23

Skills

Data Science

Machine Learning, NLP

Deep Learning

Statistics

Programming

Python Programming

Unix Scripting

Libraries

NumPy, Pandas, MatplotLib, Seaborn, Scikit-Learn, NLTK, Keras on top of Tensorflow

Databases/ SQL

Hive, Impala, MongoDB

SQL/PL SQL

MS SQL, Teradata

IDE

Jupyter Notebook

Spyder

Pycharm

Google Colab

Summary

Data Scientist with 3 years of relevant experience leveraging statistical modelling with Algorithms, data mining, machine learning algorithms, Artificial Intelligence, Machine Learning, Python with libraries such as Sklearn, Numpy, Pandas, Matplotlib, Seaborn and Tableau for Data Visualization.Total 7 years of extensive experience in various platforms like data science, ETL and BI.

Projects

- Housing Debt restructuring
- Customer Relation Sentiment Analysis
- Credit Risk Modelling
- Spam Email Detection using NLP
- Worked on Deep Learning Neural networks like ANN, CNN and RNN

Work Experience

Msys Technologies (Oct'19 - Present)

Evaluation and classification of credit for Housing Debt

This project refers the process of a lender determining whether the current evaluation of the property based on property matrices justifies with the loan re-payment schedule, otherwise proper re-structuring of the loan amount is required.Also, borrower's loan application needs to be within acceptable risk.

- Pre-process the data with more than 1 million records.
- Perform EDA on data, understand trends
- Perform Feature Engineering wherever viable
- To build and Train a Machine Learning Model using Logistics Regression, Random Forest Classification, XGBoost Classification, and using cross validation for model selection.
- To find the accuracy of the Model Prediction using Classification Reports, Confusion Matrix, ROC AUC Score.

Cloud Deployments

Deployment in Heroku

Deployment in AWS POC

Education

Degree -B.E.,

Year of Passing - 2012

University- Nagpur University

Nagpur, MH

Personal Info

DOB - 23/09/1989

Marital Status - Married

Place - Pune

Mobile - +91 96897 12560

Work Experience

L&T Infotech Limited (Nov'17 - Oct'19)

Customer Sentiment Analysis

This project refers to the prediction of whether a particular customer ceases his or her relationship with the bank based on parameters identified by the bank.

- ✧ Pre-process data from Excel, csv, databases and join/merge dataframes.
- ✧ Perform Data analysis, feature engineering and feature selection.
- ✧ Building data modelling pipeline to train the model.
- ✧ To build and Train a Machine Learning Model using Logistics Regression.
- ✧ Involved in parameter tuning process for optimal model hyper parameter optimization.
- ✧ Also worked on clustering techniques like DBSCAN and K-Means

Credit Risk Modelling

Objective of the project was to predict whether to clear the loan for the new customer or not based on the historical data of sample customers.

- ✧ Understanding the trend using Heatmap and seaborn library.
- ✧ Converting the categorical features into numerical features before model building. Feature scaling performed.
- ✧ The missing values were imputed and outliers were treated. The skewness of data was corrected using log transform.
- ✧ ML modelling was started with logistic regression with stratified K-Fold cross validation for accurate predictions.
- ✧ Involved in optimizing the hyper parameters.

Internal

- ✧ Dimensionality reduction using PCA
- ✧ Image Classification using CNN and ImageDataGenerator, ImageNet
- ✧ Worked on advanced RNN architecture like LSTM
- ✧ Also worked on VGG16 and VGG19 for multi-class classification.
- ✧ Deployment of a Spam classification model in Heroku.

Past Work Experience (ETL & BI, Data Analytics)

Cognizant Technology Solutions (Dec'15 - Nov'17)

Global Systemically Important Banks

As a regulatory response to the revealed vulnerability of the banking sector in the financial crisis of 2007/08, and attempting to come up with a solution to solve the "too big to fail" interdependence between GSIBs and the economy of sovereign states, the Financial Stability Board (FSB) started to develop a method to identify GSIBs

- ✧ Validate the source system data coming from different sources
- ✧ Validate the mappings by writing queries as per Business Requirement Document and Functional Specification Document and to check for the correctness of the data
- ✧ Created CDC and ETL scripts for business requirements

Infosys Limited (Sep'13 - Dec'15)

Nil Balance Account Closure

Objective of the project was the use of ETL and BI to identify the customers who have not transacted for a long time and to close those accounts suitable.

- ✧ Validation of Informatica and Ab-initio scripts
- ✧ Worked on Data Analysis, has experience on Unix and SQL.
- ✧ Worked on Hive and Impala
- ✧ Extensive use of SQL for validation of data.