Abdul Jameel Shaik

+917569358391 | abduljameel999@gmail.com | Hyderabad, India 500008 |
https://www.linkedin.com/in/abduljameelshaik

SUMMARY

Data Scientist with 4 plus years of Professional experience in Machine Learning and Data Analysis projects, Quality and Process Management. Currently working as a Senior Data Scientist and having end to end implementation experience in Data Science projects focusing on maximizing client satisfaction, process compliance and business value articulation, comprehensive experience in leading teams & multiple projects.

SKILLS

- Python and R programming
- Machine Learning (Supervised and Unsupervised Algorithms)
- Alteryx ETL
- Computer Vision, GAN's
- CNN, RNN, LSTM
- Tableau, Power BI
- AWS sagemaker
- SQL Queries
- Generative AI- Lang chain, LLM, Hugging Face (Basics)

WORK EXPERIENCE

Senior Associate - Data Scientist, PwC India, April 2021-Current
Hyderabad, India

Project 1: Go To Market Tool - Beverages Manufacturing Industry (May2021-July2021) - North America
Business Problem: Due to the company's large distribution network and many manufacturing plants, the client's requirement was to have a smooth transitioned workflows for each of their bottler plants and using advanced analytics techniques, cluster potential prospects and create new business opportunities
Business Objective: Using Alteryx ETL tool create workflows for each plant and push the data to Amazon work space
Business Constraints: Reducing the complexity of the workflow
Key responsibility: Design complete process workflow in **Alteryx ETL** for multiple bottlers and also write clustering code using **R programming** to segregate potential prospects based on consumption volume and frequency of orders and upload the final data to the Amazon S3 database

Project 2: Predictive Prospect Scoring Model -Beverages Manufacturing Industry (July2021-November 2021)
Business Problem: To increase the customer base and potentially target only those business who has the potential to become a prospective customer. Design and develop a **"PROSPECTIVE LEADS TOOL"** for the company.
Business Objective: To identify the potential customers and increase the market size of the outlets for North America base and Canada

Business Constraints: Random Features identification for selection of important variables from the data.
Key responsibility: Build a Machine Learning model **(XG Boost)** for USA and Canada, deploy the model in AWS Sagemaker using AWS LAMBDA and AWS GLUE build CI/CD pipeline and render the results in **PowerBI**
Project 3: Weight Variation in Sidewall and Tread-Tyre Manufacturing Industry (Client location) (Dec2021-Mar 2022)
Business Problem: As part of IR4.0 implementation in the manufacturing plants of the client in Chennai and Gujarat. Various use cases have been identified where optimisation, digitisation can be done in both plants, one of the use cases was to identify why there is **weight variation** in the side wall and thread of the tyre, and what factors are potentially affecting and what can be optimised in order to achieve tolerance level weight variation levels all the time
Business Objective: To predict and analyse the Weight variation in the side walls of various SKU's.
Business Constraints: Manual collection of data from MES and understanding and identifying the data dictionary and identifying the feature variables.
Key responsibility: Data Preparation, Data Analysis (EDA), build analytical dashboards using **Tableau** for data visualisation. Effectively used data blending, filters, parameters, calculated sets for preparing dashboards and worksheets using **Tableau.**

Project 4: Demand Forecasting-Banking Industry (Mar 2022 – Sep 2022)
Business Problem: Design and analyse potential growth in various commodities and how business loans can be provided to countries such as India, Myanmar, Nepal, Bhutan, Bangladesh in much efficient way so that bank can avoid NPE's and also provide advisory line services
Business Objective: one part of the whole project was to gain understanding on production and consumption of each country and forecast their growth for the next ten years based on which further business decisions can be taken by management
Business Constraints: Accurate data was not available as we have to merely depend on online sources.
Key Responsibility: Collecting data from different sources and preparing excel sheets and building **Demand Forecasting** model to perform forecasting on the data which is later on used by other road and transport team for further analysis. Historical data of 5-15 years is used for various commodities and its products as per HS code and **Demand Forecasting** algorithm was used to predict **Production** and **Consumption** for next 10 years for respective commodities and all countries.

Project 5: Large Scale Business Insurance Industry – North America (October 2022 – Feb 2023)
Business Problem: Setting up automated insurance quotation system for large properties in for USA client
Business Objective: Perform overhaul of existing excel based quotation system to automated system using **R programming**
Business Constraints: Understanding the existing system and adding extra policies to the code
Key Responsibility: carrying out day to day tasks of understanding the existing logic and converting them into **R scripts** and build a **R package** that can be deployed on client azure network and run test cases to check the functionality of the package.

Project 6: Financial Modelling SAP BTP 360 (April 2023- Present)
Business Problem: Creating SAP models in S4/HANA for leadership reporting

Business Objective: Working along side with Mckinsey & Co. helping them with building and testing the models
Business Constraints: Time taken to load multiple models should be reduced
Key Responsibility: I was tasked to work on loading data into all models in QA and also transform the logic used in **Alteryx** and create process flow that can be used by SAP Consultant to create models. I worked on building SAP models using calculation views and other multiple methods in **SAP BTP**. Developed advanced **Tableau dashboards** that allowed us to track wages and layoffs of employees for multiple business locations. Designed stored procedures, triggers, tables and views for several applications while generating comprehensive analytical reports by running SQL queries against current databases to conduct thorough data analyses.

Associate Consultant - Data Scientist, Innodatatics, March 2019-March 2021 Hyderabad, India

Project 1: Electric Vehicle Battery Breakdown Predictive Analytics (Dec 2020 – Feb 2021)
Client: Internal
Business Objective: Electric vehicles, the new age transportation system which is on rise in the automotive market with vehicles being sold all over the country, our client supplies batteries to these manufacturers and would like to have their huge batteries fail proof otherwise it would impact their business.
Methods and Algorithm: Data generation, attributes selection, Exploratory Data Analysis, ML algorithms such as **Artificial Neural Networks, RF, Logistic Regression, XG Boost, SVM.**
Program Layout: In this project, we started of collecting attributed based on client's information and research, then data cleaning and data mining procedures were applied to check the distribution of data and necessary steps were taken, model building using supervised and unsupervised learning algorithms were applied and accuracies measured.
Responsibilities: Data collection research work, EDA, Model Building
Business Impact: Our model can predict with **98 percent accuracy** when a EV battery could fail given conditions, our client will use it to test and improve their production of batteries.

Project 2: Predict Machine Component Failure in Food and Beverages Industry
Client: internal (July 2020 – Nov 2020)
Project Objective: In Manufacturing process of the soft drinks, right from the beginning of bottle casting to filling the drink and labelling those bottles, high level precision sensors are mounted to track the number of bottles being produced and if a component or part of a production line is stalled then it will lead to a loss of money and delay in times to fix it. So, using the sensor data we must predict when a machine will potentially break down.
Methods and Algorithm: Data Cleansing, EDA, Extraction data from Influx data base, automate EDA using **SweetViz** and Bokeh libraries of Python, applied ML algorithms such as Survival Analytics **Kaplan Meier** algorithm, Forecasting ARIMA, SARIMA and Mean Time Failure prediction.
Program Layout: In this project, we used **Bokeh** to automate our EDA process and using batch file of task manager data extraction from database is done at regular intervals, and for different components analysis by using **SARIMA** Forecasting Algorithm to predict when the next breakdown can happen.
Responsibilities: Worked on Model Building and Evaluation of the models.

Business Impact: With the implementation of the Machine Learning Algorithm, we were able to forecast the breakdown time, so that the immediate measures can be taken by the client.

Project 3: Medical Insurance and Employee Claim Analytics. (Dec 2019 – June 2020)
Client: Malaysian Insurance Company -Malaysia
Project Objective: The client wants to predict the claim amount a claimant can lodge given a disease or other illnesses.
Methods and Algorithms: Data Cleansing, Feature Selection, L1 and L2 Regression, Label Encoding.
Program Layout: In this Project, Regression Techniques such as **Lasso and Ridge Regression** were used for feature selection and Model Building, and we got an over accuracy of the model around **93%**.
Responsibilities: Worked on Data Cleaning, Data analysis and Model Implementation.
Business Impact: We were able to improve/predict the customer claims and retain customers from churning and minimizing the loss to the company.

Project 4: Wind Turbine Failure Prediction. (April 2019 – Oct 2019)
Client: Internal Project – Hyderabad, India.
Project Objective: Client manufactures, and services wind turbines and any component failed will cost to client a huge delay in changing the spare part and loss the company gets for not operating in those periods.
Methods and Algorithms: Data Collection from sensors, Mean Time Failure Analysis, EDA, Classification Algorithms used, such as Random Forest, XG Boost, Logistic Regression and SVM for Experimental Analysis.
Program Layout: Using each classification algorithm trained and test models batch wise for different components that can potentially fail.
Responsibilities: Worked on EDA, Model Building using **Random Forest, XG Boost, SVM** on Wing Motor Component Sensor Data.
Business Impact: Building a classification model which can predict a failure of component helped the client to know about it before hand from a longer distance and necessary steps could be taken whilst servicing our model accuracy was 95%.

EDUCATION AND TRAINING

Bachelors of Technology
Electronics And Communication Engineering, Jawaharlal Nehru Technological University, Hyderabad July 2013

LANGUAGES

English: First Language

English: C2

Proficient