# Mohammad Najeed Osmani *Data scientist*


najeedosmani2929@gmail.com

8317646743

Karimnagar, Telangana, India

https://www.linkedin.com/in/najeed2929/

https://github.com/najju29

https://www.kaggle.com/najeedosmani


## Skills

**Core Skills**
Business Intelligence, Machine Learning, Data Analysis, Data Analytics.

**Technical Skills**
Python, SQL, Spark, Open AI, Keras, Pytorch, Scikit learn, Faiss, Redis.

**Cloud/Deployment**
Azure, Teradata, Docker, Kubernetes, Databricks, Cognitive services, GitHub.

**Practical Skills**
Machine Learning, Natural Language Processing, MLOPs, Language Models, A/B Testing

## Professional Experience

2021/09 – present
Jaipur, India

**Data Scientist**
*Celebal Technologies*

**Customer Analytics:**

- Developed a customer analytics and recommender engine project.
- Utilized Databricks for data processing and batch pipeline execution.
- **RFM analysis:** Conducted customer-level evaluation based on lifetime monetary and frequency metrics.
- Classified customers into "High-level" and "Low-level" segments with predefined thresholds.
- Created a recommender engine to suggest personalized products to customers.
- Implemented recommendation algorithms, including collaborative filtering.
- Demonstrated proficiency in SQL for data querying and analysis.
- Leveraged PySpark for script development and data manipulation.
- Utilized Spark for distributed data processing.
- Delivered actionable insights leading to improved customer engagement and revenue growth.

**Trade Assistant Bot:**

- I have created a trading bot for a global commodities company that efficiently handles trade rules, regulations, and tariffs.
- The bot communicates with PDFs and Excel data, enabling quick and accurate access to vital information, streamlining operations, and decision-making in international trade.
- Structured(Excel) and Unstructured(PDF) data were handled separately while feeding to the model, preprocessed data that better suits the model's requirements.
- Generated OpenAI embeddings (Ada) of all the parsed documents and stored them in a Redis db(Indexer).
- Providing the user's query and relevant documents to the GPT-3.5 turbo model's prompt to extract and rephrase the required answer.
- Deployed the whole solution using Azure app services, Integrating the solution with the Front-end.

**Allocated Water-Oil-Rate Prediction:**

- The project aim was to perform a Regression model which can predict the Allocated Water-Oil-Rate.
- Exploratory data analysis was done on the data by identifying all important patterns and insights.
- The model was built on this data and we further minimized the error by applying feature engineering techniques to remove lesser important features.

- Tuning the model using Hyperparameter tuning techniques like Optuna and Bayesian Optimisation.
- Batch deployment of the model was made using Azure ML Studio and Azure Data Factory and scheduled this pipeline to run on an hourly basis.

2021/05 – 2021/08
Jaipur, India

**Data scientist - Trainee**
*Celebal Technologies*
**Customer churn prediction:**

- The goal of this project was to predict and identify the customers who are at higher risk of churning out based on historic data.
- Exploratory data analysis based on historic sales data for analyzing and finding out patterns and insights.
- Predicting a set of customers who are at a high risk of churning out, with predictive analytics machine learning models based on historic sales data.
- Data Ingestion Pipeline from Azure Synapse and running ETL Job over Databricks.
- MLOPs is achieved by deploying the model on Azure app services, versioning code on Azure Repos which does CI-CD, and Monitoring the Model using Azure Monitor.

## Achievements

**Kaggle Bronze Medal**
*Won Bronze medal in a Kaggle competition "GoDaddy - Microbusiness Density Forecasting". The goal of this competition was to predict monthly microbusiness density in a given area. I have developed an accurate model trained on U.S. county-level data.*

**Runner Up (Rank 2) in a ML Hackathon**
*In this competition provided the visitor's session data, we were challenged to come up with a regression algorithm to predict the time a buyer will spend on the platform.*

**MachineHack Best Global Rank - 42**
*I have particpated in more than 60+ Hackathons and Competitons collectively and was ranked in top 1 percentile in more than 40+ hackathons, making me exposed to versatile problem statements and applying unique solutions to solve them up.*

## Certificates

**Databricks Certified Machine Learning Professional**
Earners of the Machine Learning Professional certification have demonstrated an ability to perform advanced machine learning tasks using Databricks Machine Learning and its capabilities.

**Academy Accreditation - Databricks Lakehouse Fundamentals**
Earners of the Lakehouse Fundamentals accreditation have demonstrated the understanding of fundamental concepts related to Databricks Lakehouse Platform.