

SUSHANT SUR

Data Scientist

My Contact

✉ sushant_sur23@yahoo.co.in

☎ +91 966-350-1146

📍 Varthur, Bangalore

🐙 github.com/sushantsur23

in linkedin.com/in/sushant-sur-b1b98319/

Skills

- Python, VS Code, Pycharm
- Matplotlib, seaborn, scikit learn, numpy, pandas
- Mysql, Mongo DB
- Tableau, Power BI
- Machine Learning, Docker, AWS, Hive
- Classification, Regression, Clustering
- Github Actions, Circle CI, Git bash
- ML Flow, DVC, CICD pipeline, Linux, Evidently AI, S3, Vault, Hive, Impala, Data Governance, ETL pipelines
- NLP, Hugging Face, POS Tagging, (SST, TTS)
- Speech synthesis with IBM Watson API
- Pytorch, Tensorflow

Education Background

- Liverpool John Moore University
MS in ML And AI
2024-2025
- IIIT Bangalore
PG Program in ML And AI
2023-2024
- ISBM, Kolkata
MBA Finance
2008-2010

Certifications

Scrum Master, Data Science
Optimize data pipeline with Azure Data Factory and Synapse, Build data pipeline with Azure data Factory (Pluralsight)

Hobbies

Swimming Travelling, Gym, Technology

About Me

Data Science professional with statistical analytics capabilities, well versed with data cleaning, data transformation techniques to provide data-driven recommendations and an active proactive problem solving skills. Possess diverse experience in executing projects in validating techniques and gathering requirements ensuring business growth and development.

Professional Experience

Dell (BCT Consulting)

11/2022 – 01/2023

Key responsibilities:

- Supported data migration projects where moving the files from source to production and scheduling jobs with Control-M.
- Managed CICD pipelines in gitlab as per the devops standards with the stories updated on Azure devops dashboard spread across various sprints and using the pipelines through various scans/vulnerability checks to move files in hadoop nodes.
- Updating the SQL scripts as per the stories updated from the client side and sending completed stories for production once in a month.

Travelopia. | Financial Planning Data Analyst

08/2021 – 04/2022

Key responsibilities:

- Validating the transactions through data received from bank server, preprocessing the data and keeping only the important features needed.
- Creating data ingestion and transformation pipelines, finally using various classification algorithms with ML Flow architecture.
- Finally deploying it on AWS SageMaker with continuous deployment pipeline. Model accuracy was around 83% and a continuous retraining approach was used with any new data set received. Additionally scheduling the task with Airflow.

EY | Lead Finance Analyst

03/2020 – 12/2020

Key responsibilities:

- Finding out customer churn rate would be needed in any business in order to increase customer engagement or growth of sales for the organization. Here we tracked data drift with Evidently AI.
- We achieved an accuracy of 83.5% on the trained model - Used data ingestion and data validation pipeline before passing it to
- classification algorithms. Used Kafka Confluent API's to get the IOT data and transfer it to Mongo DB for storing, Circle CI helped to deployment pipeline, finally created a docker image on AWS ECR and deployed on EC2.
- Reading the IOT data from telemetric device through kafka and then preprocessing it in Pyspark taking care of the huge volume. Then finally applying various classification algorithms to predict the score of the respective driver which was helpful to decide the insurance premium amount.

Accenture | Senior Analyst

05/2015 – 12/2019

Key responsibilities:

- Analyzing reasons for variance across contracts with advising on the profitability improvement plans and creating tableau dashboards to show insights/impact of various KPI's on the contract financials.
- Visualization included cross tabs, bar & line charts combined together and waterfalls for different LOBs in Tableau.
- Row-level security was implemented in the Tableau dashboard so that each line of business can only see their respective data while the leadership group can see all the data of the full business.
- Extracting data through SQL queries and preparing various financial reports for the stakeholders.
- Monitoring and managing ETL pipelines in SSIS.

Capgemini | Senior Process Associate

07/2011 – 05/2015

Key responsibilities:

- Reconciling cash balance with books vs bank statement, Analyzing variance for various GL accounts and posting corrections entries wherever needed. Handled remote transition for Techops.

Projects Undertaken

NER implementation with XLM Roberta

Key responsibilities:

- Many commercial use cases can be resolved with the application of the NER principle, however there is still much room for improvement. This will occur in many NLP use cases in order to either swiftly filter documents or fast extract key points from a vast corpus, or it could enhance the system with recommendations.
- With the xtreme dataset, training F1 score came out to be as 80% and test F1 score came as 82%. NER tags inorder to tag information related to person, organization and location were used in the project.

Chatbot with Dialogflow

Key responsibilities:

Created a Google DialogFlow chatbot which can predict live weather conditions for some of the famous cities in India and UK, additionally it can also respond to some jokes. Integrated with a telegram app with terminator23@bot.