

Sumith Singh Kothwal

Tel: +33 753179873 | Email: sumithsingh27@gmail.com | [LinkedIn](#)

PROJECTS

AI-Powered Cancer Detection Platform:

- Created a custom dataset of blood smear images from multiple sources, published on Kaggle for community use. Implemented self-supervised learning using TensorFlow, achieving 95% accuracy after fine-tuning on labeled data.
- Developed a comprehensive healthcare platform integrating the ML model with a user-friendly interface for doctors and patients for patient data analysis. Built an AI assistant (Dr.Clara) using fine-tuned LLaMA-2 for 24/7 patient support and automated report generation.

Price Prediction System:

- Architected and deployed an end-to-end ML solution using Python and Linear Regression, achieving 93% prediction accuracy by analyzing historical price data and market indicators. Built an interactive UI with Streamlit for real-time price predictions and trend analysis.
- Designed a robust data pipeline using Airflow DAGs for hourly data ingestion, implemented data quality checks using Great Expectations, and monitored model drift through Grafana dashboards. Used FastAPI to serve predictions with 99% uptime.
- Created a scalable architecture using PostgreSQL and SQLAlchemy for data storage, Docker for containerization, and AWS EC2 for deployment. Set up automated CI/CD using GitHub Actions for seamless updates.

Sentiment Analysis System:

- Built a comparative study of sentiment models (BERT, RoBERTa, BiLSTM) reaching 91% accuracy on tweet classification. Created an extensive analysis report comparing model performances across different metrics and computational requirements.
- Developed an automated text processing pipeline using NLTK and spaCy, handling different tweets. Built an interactive dashboard using Streamlit to showcase real-time sentiment analysis and trend visualization.

Healthcare Data Management System:

- Led a team project to design a multi-database healthcare system using MongoDB for patient records, Neo4j for relationship mapping, and Redis for caching. Reduced query response time by 40% through optimized database design.
- Created RESTful APIs using Flask and FastAPI to handle medical data operations, implementing comprehensive unit testing with pytest. Built a secure authentication system for patient data protection.

Customer Churn Analytics:

- Built a churn prediction model using Dataiku's AutoML capabilities, analyzing customer patterns and identifying key indicators through feature importance analysis.
- Created interactive dashboards using Dataiku's visualization tools to present model insights and top influencing features, providing business teams with actionable recommendations for customer retention strategies.

PROFESSIONAL EXPERIENCE

White Turtle IT Technologies, Hyderabad, India

Junior Data Engineer

July 2022 - October 2023

- Built data pipelines using Python and SQL to process e-commerce data from PostgreSQL databases, working with senior engineers to learn ETL best practices and implement data validation checks that helped identify data quality issues.
- Created Python scripts using pandas and FastAPI to automate daily data extraction tasks, handling product and inventory data synchronization between systems. Documented processes in Confluence for team reference.
- Gained hands-on experience with Git version control and Jira for task tracking while participating in daily standups and sprint planning. Learned to write clear technical documentation and communicate data issues effectively with business teams.

EDUCATION

- EPITA, Paris, France - MSc Data Science and Analytics

February 2024 - Present

- Vardhaman College of Engineering, Hyderabad, India

July 2018 - May 2022

- B.E Electrical and Electronics Engineering, Graduation mark: 8.19/10 (Distinction)

ADDITIONAL INFORMATION

Technical Skills: Advanced proficiency in Python and SQL (Certified) with strong knowledge of database technologies including PostgreSQL, NoSQL, and MongoDB. Extensive experience with Python Data Science stack (NumPy, Pandas, scikit-learn, Matplotlib, Seaborn, TensorFlow, Plotly). Skilled in cloud services using AWS (EC2, S3, Lambda) and version control with Git. Proficient in data visualization tools including Tableau, PowerBI, Dataiku (Certified), and Advanced Excel. Experience with DevOps & MLOps tools such as Docker, Airflow, and MLflow. Strong foundation in statistical analysis methods including regression models, statistical modeling hypothesis testing, data validation & quality assurance, and time series analysis.

Languages: Hindi (Native), English (Fluent), French (Intermediate)

Interests:

- Passionate about artificial intelligence research, particularly focusing on Deep Learning papers and advancements in the field. Regular badminton player who enjoys competitive matches.
- Enthusiastic traveler who loves exploring new destinations and experiencing different cultures through local cuisines and adventures.