KAMESH KOTWANI

LinkedIn | +447932851567 | Portfolio | kameshkotwani@gmail.com | GitHub | Medium

SUMMARY

A Data Scientist with 2+ years of experience in data world with 85% accuracy for language models, automating manual tasks with 30% operational efficiency, and leading end-to-end processes for stakeholders. Excellent record in boosting customer retention, accelerating decision-making by 25%, and contributing to revenue growth. Proficient in a suite of advance data science tools, specializing in machine learning, deep learning and Machine Learning Operations.

SKILLS

Languages & Databases	Python, SQL, R, Scala, Java, Oracle Apex, MySQL, Postgres, NoSQL, Elasticsearch.
AI Models & Statistics	NLP, DVC (Data Version Control), Predictive Modelling, Regression Analysis, XGBoost, Statistical Modelling, Probability Theory, Calculus, Large Language Models, Generative AI.
Analytics Tools	Power BI, Tableau, KNIME Analytics, IBM SPSS.
Libraries & Frameworks	Matplotlib, Seaborn, Plotly, Pandas, NumPy, SciPy, Scikit-learn, TensorFlow, Pytorch, Docker, Keras, Flask, Hadoop, Airflow, Spark/PySpark, NLTK, LSTM, SpaCy, GitLab, Git.
API & Cloud Computing	MLflow, AWS, Azure, GCP, CI/CD Pipelines.

EXPERIENCE

IQVIA | Big Data Engineer | *Bangalore, India* *June 2022 - August 2023*

- Collaborated with a team of 10+ members, optimizing Spark and Airflow pipelines and reducing modelling and run time by 30%.
- Led end to end data processes for 20+ pharmaceutical stakeholders, boosting customer retention by 12% and decision making by 20%.
- Enhanced efficiency in team workflow and automated redundant tasks using bash and Python, which increased operational boost by 25%.
- Authored Scala code on Apache Spark to optimize and decrease data processing pipeline runtime from 10-12 hours to 6-7 hours for intensive containing terabytes of data.
- Programmed Airflow pipelines for cluster cleanup automation, which helped to break the limited space deadlock using Airflow. It shortened the low space error debugging issue by 90% and was highly appreciated by the team.

CDAC | NLP/ML Project Engineer | *Pune, India* *October 2021 - June 2022*

- Facilitated the kanthasth project by CDAC, involving the development of an AI chatbot using JS/NLP (https://kanthasth-rajbhasha.gov.in/), which lessens manual query from users by 35%.
- Worked with a 5+ member team to create an NLP and text analytics dashboard for research purposes, reducing analytics time by 50%.
- Created Language Translation Models across various languages with an accuracy of 85% using Transformers and Deep Learning.
- Crafted various automation scripts with bash and Python to increase team workflow, which decreased manual tasks by 80%.
- Engineered and implemented frontend user interfaces for NLP initiatives using Django and Flask, enhancing user accessibility and leading to a 25% increase in user retention rates and a 20% boost in user satisfaction scores.

PROJECTS

Movie Recommender System **GitHub Link**

- Created a movie recommender system using collaborative filtering process which recommends similar movies using cosine similarity distance among 5000 movies.
- Developed cutting-edge data wrangling strategies to combat imbalances, utilizing synthetic data augmentation and data mining techniques for insightful Exploratory Data Analysis. Facilitated user-friendly access to solutions through deployment on streamlit platform.

WhatsApp Chat Analyzer **GitHub Link**

- Produced a sophisticated NLP tool leveraging libraries such as spaCy, scikit-learn, and WordCloud, enabling sentiment analysis and text mining capabilities to derive valuable business intelligence from WhatsApp chat data.
- It analyses chats and gives interesting insights about user chats such as most active time, frequently recycled words and contacts.

Book Recommender System **GitHub Link**

- Designed a book recommender system using content filtering which recommends books and shows top 48 books.
- Indulged in Data Cleaning, Data Analysis, kNN Imputation for missing values, Feature Engineering and Feature Scaling to increase models' performance and accuracy.

Leaf Disease Detection using Deep Learning CNN

- This is the university research project which using CNN architecture to detect if the leaf has a certain disease or not with 78% accuracy.
- Many CNN architectures adopted to create and test models such as AlexNet, ResNet, ImageNet. Optimizers such as Adam, Adagrad also experimented to train faster.
- Trained in multiple iterations using MLOps, DVC (Data Version Control), CI/CD, sklearn-pipelines, Keras and PyTorch backend.

EDUCATION

- Masters in data science | University of Sheffield | *Sheffield, United Kingdom* *September 2023 – September 2024*
- PG Diploma- Big Data Analytics | CDAC | *Pune, India (Distinction)* *May 2021 – October 2021*
- Bachelors in computer science | RGPV | *Jabalpur, India (Distinction)* *August 2016 – August 2020*