Rishi Sharma
Data Scientist

Mobile: 8860551924
Portfolio Website: https://rishi2628.github.io/Resume_Rishi/

SKILL SET

Programming Languages and Databases
• **Programming Languages**: Python, SQL, Git
• **Automation Tools**: Selenium
• **SQL Databases**: PostgreSQL

Data Analysis and Modelling
• **Predictive Modelling**: Regression, Random Forest, XGBoost, Recommendation Systems, Kmeans, Clustering
• **Data Analysis Tools**: Pandas, NumPy, SciPy
• **Visualization Tools**: Matplotlib, Seaborn, Plotly

Development and API Tools
• **Development Tools**: Jupyter Notebook , PyCharm , Databricks
• **API Tools**: Postman, FastAPI

Machine Learning and AI
• **Libraries**: Scikit-Learn, Keras, TensorFlow
• **AI/ML**: Neural Networks , Deep Learning , Large Language Models (LLM) , Natural Language Processing (NLP)
• **AI and ML Services**: Langchain, OpenAI API , GPT , GPT-4, Prompt Engineering
• **Serverless**: AWS Lambda, Amazon Web Services

EXPERIENCE (8 + YEARS)

Machine Learning Engineer | Deloitte, Gurgaon (2021 to Present)

Project 1 – Options Trading Algorithm
- Developed machine learning algorithms to predict optimal option trades by analysing complex market data and volatility trends.
- Built end-to-end data pipelines to scrap, collect, clean and transform large datasets including underlying price action, options chain data (strikes/expiries), implied volatility, greeks, volume and open interest and macro/market indicators.
- Built end-to-end data pipelines to engineer robust feature sets using historical price data, trading volume and options chain metrics, specifically focusing on the "Greeks" to improve decision accuracy.
- Engineered predictive features such as volatility surfaces, term structure signals, momentum/mean-reversion indicators, liquidity filters and event-driven flags to improve signal quality and reduce noise.
- Implemented a back testing framework to validate strategies against historical data and deployed the solution on AWS for scalability, utilizing services like EC2 and S3.

Project 2 – Credit Risk Modelling – Loan Defaults
- Developed machine learning models to classify high-risk credit defaulters with no prior loan history.
- Engineered data models using user data such as bank transactions, employment details, and asset information.
- Utilized user-specific data including age, income, occupation, and collateral to predict key metrics such as Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD).
- Implemented backend services in Python to process and analyse large datasets efficiently and deployed the application on AWS utilizing services like EC2 and S3 to ensure scalability, reliability, and secure hosting.
- Achieved a 3% reduction in default rate, saving the client approximately $50K across seven bank branches through effective risk modelling.

Project 3 – LLM Based Audit Report Generator
- Developed an audit summarization tool to generate comprehensive reports from client document files, enhancing data insights and report summaries.
- Engineered the backend of the application using Python, FastAPI, and AsyncIO for efficient and scalable performance.
- Implemented advanced techniques such as MapReduce, Retrieval-Augmented Generation (RAG), and vector databases in Langchain to optimize operations and minimize latency.
- Integrated OpenAI models (e.g., GPT-3.5, GPT-4, Davinci) with Langchain for sophisticated summarization and categorization tasks.
- Deployed the application on AWS, utilizing services like EC2 and S3 to ensure scalable, reliable, and secure hosting.

Data Scientist | Optum (UnitedHealth Group), Gurgaon (2019 to 2021)

Project 1– NLP Based Service Navigation
- Creating machine learning models to perform classification for the user submitted UHG insurance claims and paperwork.
- Generating word embeddings using libraries such as Gensim word2vec, while utilising transformers, spacy, flair, Huggingface - GPT, BERT
- Segmenting user documents using OpenCV and converting image to text using OCR Engine like Tesseract and Amazon Textract.
- Integrating the model in production pipeline over AWS. Cutting down operational cost by $ 100K dollars across three claim centres.

Project 2– Insurance Recommendation Model
- Building Machine Learning models for health care providers to shortlist customers for better sales & lower customer acquisition cost.
- Analysing RX and Medical claims history of members, from hive database along with other related details like Demographics,
- Online activity, AWV and other factors that could aid in the improvement of acceptance rate.
- Creating classification models based on the above user data. Recommendation from model led to 4% volume uplift. Received recognition from client's end.

CERTIFICATIONS

AWS Certified Machine Learning Speciality
Validation Number: 48365a43d996414db0cbeb053211fb21
Issue Date: November 15, 2024
Expiration Date: November 15, 2027