Rishi Sharma
Data Scientist

Mob: 8860551924
https://rishi2628.github.io/Resume_Rishi/

BRIEF

Data science professional with over 7 years of experience in predictive analysis, software development and data modeling using Python. Proficient in problem analysis and solving, with excellent communication skills for conveying complex ideas. Experienced in working with clients from diverse geographies and domains to drive better decision-making through actionable insights. An independent learner with strong teamwork ethics and a confident, articulate communicator.

SKILL SET

Programming Languages and Tools
- **Programming Languages**: Python , SQL , Pyspark , HTML , Bash , Git
- **Version Control**: GitHub Actions, Bitbucket
- **Automation Tools**: Selenium

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

Machine Learning and AI
- **Libraries**: Scikit-Learn , Keras , PyTorch , TensorFlow
- **AutoML** Tools: H2O.ai , DataRobot , Google AutoML

- **AI/ML**: Neural Networks , Deep Learning , Large Language Models (LLM) , Computer Vision , Natural Language Processing (NLP) , Optical Character Recognition (OCR) , Object Detection / Segmentation

Databases
- **SQL Databases**: PostgreSQL , Apache Hive
- **NoSQL Databases**: AWS DynamoDB , Vector Databases , TimescaleDB, Amazon Neptune
- **Data Warehousing**: Amazon Redshift, Google BigQuery, Snowflake

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

Cloud Services
- **Cloud Platforms**: Amazon Web Services (AWS) , Google Cloud Platform (GCP)
- **Containerization and Orchestration**: Docker
- **AI and ML Services**: Langchain , OpenAI API , GPT , GPT-4 , Davinci, Prompt Engineering , RAG
- **Serverless**: AWS Lambda , Google Cloud Functions

Web Development
- **Frameworks**: Django , Flask

Big Data and Data Engineering
- **Big Data Tools:** Apache Spark, Apache Kafka, Hadoop

EXPERIENCE (7 + YEARS)

Machine Learning Consultant | Deloitte, Gurgaon (3.5 + Years)
Project 1 – LLM Based Audit Summarizer
- 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.

Project 2 – Chabot with Integrated Recommendation System
- Developed a voice-enabled chatbot using OpenAI models (GPT-3.5) to take orders from customers in drive-throughs for a global restaurant chain.
- Designed and implemented the chatbot flow using Python and FastAPI, including handling edge cases to ensure robust and seamless interaction.
- Trained and integrated a recommendation system to suggest and upsell menu items, enhancing order processing speed and improving user experience.
- Engineered backend services to manage interactions between chatbot & recommendation system, ensuring efficient data processing & API communication.
- Achieved a 7% increase in average order value through model recommendations, significantly improving client service delivery and customer satisfaction.

Project 3 – 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.

Data Scientist | Optum (UnitedHealth Group), Gurgaon/ Bangalore (2 Years)
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.

EDUCATION

Maharaja Agrasen Institute of Technology, Delhi **Aug 2013 – June 2017**
Guru Gobind Singh Indraprastha University (GGSIPU), Delhi
B. Tech (Computer Science) **(Aggregate Percentage – 75%)**