**Krish Sharma**

**San Jose, CA**

**PROFESSIONAL SUMMARY**

Experienced Machine Learning Engineer/ Data Scientist with over 7 years in the industry, specializing in AI/ML solutions using GCP, Azure and AWS technologies for the past 9+ years, and over one year with Generative AI. Proven track record of deploying, consuming, and fine-tuning NLP model & LLMs such as Azure Open AI, Llama 2/3, and Hugging Face. Extensive expertise in designing and implementing scalable AI solutions. Strong proficiency in Python, TensorFlow, PyTorch, and scikit-learn, coupled with hands-on experience in deploying AI solutions on cloud platforms like Azure and AWS. Exceptional communication and leadership skills with a history of successful collaboration across cross-functional teams.

* Experience in using various packages in R and python - like NLP, pandas, NumPy, Seaborn, SciPy, Matplotlib, sci-kit-learn, Beautiful Soup, Keras, PyTorch, and TensorFlow.
* Experience with data storage and management on AWS, including creating and maintaining data pipelines and ETL processes.
* Extensive ETL testing experience using Informatica (Power Centre/ Power Mart) (Designer, Workflow Manager, Workflow Monitor and Server Manager).
* Working experience with advanced Microsoft Excel functions, ETL (Extract, Transform and Load) of data into a data mart and Business Intelligence (BI) tools like Microsoft Power BI and Tableau (Data visualization and Analytics).
* Worked on Web Applications in Azure and Azure functions to pull data from API to blob Storage and SQL.
* Worked on deploying websites in Azure and used MVC framework on the backend. Experienced in Developing PL/SQL - Procedures, Functions, SQL Scripts and Database Triggers to populate the data by applying business logic. Extensive knowledge in developing PL/SQL codes using Oracle 11g.
* Worked on Azure SQL data warehouse and database development. Experienced user of PL/SQL for development of server-end program units as well as reusable codes using Procedures, and functions and worked on ad hoc change tickets.
* Experienced with creating data-driven dashboards using Tableau and Superset to provide valuable insights into the management.
* Experienced in using Vertex AI's integration with TensorFlow Extended (TFX) for creating robust and scalable ML pipelines.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| ***AI/ML Solutions*** | Azure Open AI, Llama 2/3, Mixtral |
| ***Cloud Platforms*** | AWS, GCP, Azure AI services (Azure AI Search, Azure Open AI) |
| ***Programming Languages*** | Python, PySpark, C++ |
| ***ML Framework*** | TensorFlow, PyTorch, Langchain & Llama Index, scikit-learn, NLTK, Spacy, Pandas, NumPy, Hugging Face Transformers |
| ***Data Management*** | Vector Databases, embeddings, SQL, NoSQL |
| ***AI/ML Ops Practices*** | Model Monitoring, optimization & deployment, fine-tuning |
| ***Software Version Control & Documentation*** | Git, JIRA, Confluence |
| ***Containerization & Orchestration Tools*** | Docker, Kubernetes, AirFlow & MLFlow |
| ***Monitoring Tools*** | Power BI & Tableau |
| ***Soft Skills*** | Excellent communications, leadership, collaboration, Project management |

**PROFESSIONAL EXPERIENCE**

**Cymer, Remote April 2022 – Present**

**AI/ML Engineer (GenAI)**

* Led the design and implementation of AI models and algorithms to address complex business challenges, resulting in a 20% improvement in operational efficiency.
* Collaborated with product managers to identify AI opportunities and integrated AI solutions into existing products, enhancing overall user experience.
* Developed chatbot to summarize various documents within Cymer by applying RAG technique using Langchain, LlamaIndex and vector database (CosmosDB).
* Demonstrated expertise in AI-specific utilities, including proficiency in ChatGPT, Hugging Face Transformers, and associated data analysis methods, highlighting a comprehensive understanding of advanced artificial intelligence tools and techniques.
* Expertise knowledge on AI/ML application lifecycles and workflows from data ingestion to model deployment in cloud environment like Azure, AWS & GCP.
* Deployed and fine-tuned LLM models including Azure Open AI and Llama 2/3 to create a chatbot to find relevant content from organization documents (process documents). This chatbot improved processes and reduced content search time with better summary.
* Mentor Data Scientist/ ML Engineer to get up to speed to start contributing into the client project.
* Participate in customer communication to discuss challenges and provide development status update.
* Collaborate with cross functional team to gather requirements and explore AI solutions for their problems by developing proof of concepts (PoC) utilizing latest technologies like Azure open AI and Azure Search.

**Morgan Stanley, Pleasanton, CA Jun 2020-Mar 2022**

**Sr. ML Engineer**

**Responsibilities:**

* Utilized Pandas and NumPy for data cleaning, feature engineering, and normalization to prepare datasets for modeling.
* Designed and implemented predictive models using TensorFlow and PyTorch, experimenting with various deep learning architectures including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)to handle sequential data effectively.
* Used Python to create Statistical algorithms involving Multivariate Regression, Linear Regression, Logistic Regression, PCA, Random Forest models, Decision trees, Support Vector Machine for estimating the risks of welfare dependency.
* Led the design and implementation of a customer segmentation project using AWS S3 for data storage, Python, and Pandas for data manipulation, applying K-means clustering in Scikit-learn to segment customers, enhancing marketing strategies.
* Developed a GAN-based model to generate high-quality synthetic images for training a computer vision system, significantly improving its accuracy and robustness.
* Applied advanced natural language processing (NLP) methodologies to extract insights from unstructured data sources.
* Integrate AI/ML Model and APIs into production AWS SageMaker.
* Designed and developed natural language processing (NLP) pipelines to enhance search relevance and user experience by integrating semantic search capabilities.
* Worked with cross-functional teams (including data engineer team) to extract data and rapidly execute from MongoDB through MongD3 connector for Hadoop.
* Conducted performance testing and benchmarking of cognitive search systems to identify bottlenecks and optimize system scalability and response times.
* Dealt with large amount of cloud data storage to Identify faces of same person from Image data storage and faces with similar features using NumPy, Seaborn, PIL, matplotlib, Pandas, OpenCV and Sci-kit learn Libraries.
* Create a Flask API to process input failure log files, generate summarized content, and integrate this with a Large Language Model (LLM) to produce concise text summaries.
* Developed the different Python workflows triggered by events from other systems. Collected, analyzed, and interpreted the raw data from various clients’ REST APIs.
* Created interactive dashboards in Tableau that provide a high-level overview of transaction activities and fraud detection metrics. Used Tableau’s built-in statistical tools to perform analyses like correlation studies, regression analysis, or time-series forecasting.

**Environment**: Python, R, Tableau, Power BI, Machine Learning (Scikit-Learn, Keras, PyTorch), Generative AI. Deep Learning, Natural Language Processing, Cognitive Search, Data Analysis (Pandas, NumPy), Vertex AI, SQL, NoSQL (MySQL, PostgreSQL), Django Web Framework, HTML, XHTML, AJAX, CSS, JavaScript, XML, JSON, Flask, Agile Methodologies, SCRUM Process

**Gap Inc, San Francisco, CA Jun 2018 - May 2020**

**Data Analyst/ML Engineer**

**Responsibilities:**

• Performed data pre-processing and feature engineering for further predictive analytics using Python Pandas

• Worked with NLTK library to NLP data processing and find the patterns.

• Addressed overfitting by implementing algorithm regularization methods like L2 and L1.

• Evaluated models using Cross Validation, Log loss function, ROC curves and used AUC for feature selection and elastic technologies like Elastic Search, Kibana etc.

• Worked on Power BI dashboards and ad hoc DAX queries for Power BI. Extensive back-end support for many various reports on Tableau and Power BI.

• Implemented application of various machine learning algorithms and statistical modeling like Decision Tree, Text Analytics, Sentiment Analysis, Naive Bayes, Logistic Regression and Linear Regression using Python to determine the accuracy rate of each model.

• Maintained a deep understanding of the latest advancements in MLflow and contributed to the community through documentation, code contributions, and knowledge sharing.

• Performed data cleaning and feature selection using MLlib package in PySpark and working with deep learning frameworks such as Caffe, Keras etc.

• Participated in Business Analysis, talking to business Users and determining the entities and attributes for the Data Model. Worked with different team clients, analytics, data sources, QA and visualization.

• Worked in the agile process having expertise in the life cycle of a project from start to end.

• Created data warehouse and data marts from scratch based on data structures and reporting needs.

• Created store procedures, functions, triggers, indexes, and views based on data processing and data cleansing.

• Created various SSIS packages for data movements and transformations. Worked on Python code to load data from Google buckets to big query tables, used data flow for ELT (Extract Load Transform) in Big Query.

**Environment:** Power BI, XGBoost, ETL, Tableau, Numpy, Scipy, NLTK, PLSQL, Mlib, Mflow, Python.

**EDUCATION**

Bachelor of Technology in Vellore Institute of Technology, India