SAINADH CHEBROLU

Sainadhc292000@gmail.com || 832-396-3007 || LinkedIn ||Portfolio

EDUCATION:

Bachelor of Science – Computer Science **APR 2017 to MAY 2021**
KKR and KSR Institute of Technology and sciences, Andhra Pradesh, India.

Master of science – Engineering Data Science **JAN 2022 to MAY 2023**
University of Houston, Texas, USA.

PROFESSIONAL EXPERIENCE

Walmart Global Tech – Bentonville, AR **JAN 2024 to PRESENT**
Data Scientist – Contract

- Collaborated with the team of developers to classify the severity of the claims using predictive modeling techniques.
- Performed data quality assessments to check for mismatched data types, mixed values in the data.
- Implemented **feature engineering** by calculating the correlation coefficients using Pearson's correlation, spearman's correlation, random forest correlation.
- Resolved class imbalance in the dataset by using over sampling, under sampling and SMOTE approaches.
- Implemented cross validation techniques like K-fold, stratified to choose the best hyperparameters for the models.
- Fine-tuning the **BERT** and implementing a **CNN** over it to capture the patterns and detect the severity of the claim in the early stages.
- Calculated the various evaluation metrics such as **Accuracy, Precision, Recall, F1-score** to estimate the performance of the models.
- Developing a **claim-easy-pro** a summarizer to generate the summary using **Azure OpenAI.**
- Developed various **chunking** strategies to generate the **embeddings** and store in the **Vector DB**.
- Implemented a **RAG** pattern using **milvus DB** to generate more meaningful summaries.
- Used **Pandas, Langchain, OpenAI, Azure OpenAI** in Python for developing financial bot.
- Implemented to generate **6 Line P&L summaries** by comparing KPI' and drivers.
- Good knowledge in different prompt techniques like **one-shot, zero-shot, few-shot, Chain of thought prompting** learning to guide the model to generate the text based on the user input.
- Designed **Prompt Templates** for effective prompting and guiding the language model to generate the desired data outputs.

MoxieIT – Bentonville, AR **AUG 2023 to DEC 2023**
Jr Data Scientist

- Participated in all phases of **data acquisition, data cleaning, developing models, validation**, and **visualization** to deliver data science solutions.
- Used **Pandas, NumPy, Seaborn, Matplotlib, TensorFlow, PyTorch, Scikit-learn** in Python for developing various machine learning models.
- Developed Generative AI **LLM's** and **chatbots** with diverse functionalities, including Q&A, language translation, and document search/summarization capabilities.
- Employed **Langchain** to generate vector embeddings for the documents and Hugging Face models for querying the data sources.
- Implement **RAG** pattern to extract the context from data store and include the context in the prompt for better results.
- Developed **NLP pipelines** for sentiment analysis, including text preprocessing, feature extraction, and model training, resulting in highly accurate sentiment classification.
- Experience with NLP libraries like **spacy** and Regular expressions(regex) to perform **tokenization, stemming, and lemmatization** techniques to prepare textual data for sentiment analysis, ensuring data quality and consistency.
- Containerized the application using **Docker** and Deploy it on Azure **Kubernetes**.

SWAGG Technologies – Bangalore, India **JAN 2021 to DEC 2021**
Jr Data Scientist

- Applied classification and regression, comparing various initial models and presenting reports to other teams within the company.
- Employed **Azure Machine Learning (AML)** for training, testing, Validation, deploying, and managing machine learning models.
- Utilized AML's automated machine learning (**AutoML**) capabilities to quickly generate and evaluate machine learning models.
- Developed a predictive model to generate **risk score** for shipment at the billing time for efficient handling throughout its life cycle.
- Used **PyTorch, TensorFlow** in Python for developing various machine learning models and Artificial intelligence models utilized algorithms such as Decision Trees, Logistic regression, Linear Regression.
- Worked on **freight claims analysis** on shipments using the history of transactions with supervised learning methods.
- Used various metrics (**RMSE, MAE, F-Score, ROC, and AUC**) for evaluation and validation of the performance of a model.

Graduate Academic Experience **JAN 2022 to MAY 2023**

- Implemented CNN model to classify the hand - written digits and utilized YOLO v5 to identify the objects in the images.
- Acquired a LSTM, with Auto-Regressive model to predict the stock prices. The context-window is used to add the previous prediction to the data. The MAPE score is used to calculate the model accuracy.

- Developed Random Forest, Decision Trees, Ensemble models to predict whether the customer will pick up the drug or not.
- Utilized various hyperparameters tuning techniques Like k-fold cross validation, LOOCV to identify the best subset of parameters.
- The model accuracy is validated using various metrics like Precision, Recall, F1-score, Accuracy.

CERTIFICATIONS:

- Microsoft certified Azure Data scientist Associate
- Infosys certified python programmer

SOFTWARE PROFICIENCY:

Programming Languages: Python, C, Java.
Machine Learning: Numpy, Pandas, TensorFlow, Matplotlib, Seaborn, Scikit-learn.
Generative AI: Large Language Models, Transformers, Embeddings, Prompt engineering, Langchain.
Databases: MySQL, Oracle, MongoDB, Milvus.
SDLC Methodologies: Agile, Scrum.
Project Management: JIRA, GIT.
Cloud: AZURE, GCP