**Mohamed Yusuf N**

**Data Scientist**

KEY SKILLS

PROFESSIONAL PROFILE

* Data Science Expertise: 4+ years of experience in Machine Learning (ML) & Deep Learning (DL), specializing in predictive modeling and Transformer-based generative models with cloud deployment.
* Technical Proficiency: Strong expertise in PyTorch, TensorFlow, AWS, Azure, Python, SQL and advanced ML/DL frameworks.
* Machine Learning & Deep Learning Projects: Hands-on experience in developing and optimizing ML & DL models, Transformer Based Model, NLP, Computer Vision, and Generative AI.
* Data Storytelling & Visualization: Skilled in presenting complex data insights using Matplotlib, Seaborn, Plotly ensuring data-driven decision-making.

TECHNICAL EXPERTISE

TECHNICAL EXPERTISE

|  |
| --- |
| **Languages**: Python, SQL |
| **AI Framework :** PyTorch, Scikit-learn, HuggingFace,TensorFlow |
| **Deployment :** Flask, FastAPI, Streamlit |
| **Databases:** MySQL, PostgreSQL |
| **Cloud Platforms :** AWS, Azure |
| **Data Visualization**: Tableau, Matplotlib, Seaborn |

EDUCATION

QUALIFICATIONS

B.Sc (Mathematics) – Sadakathullah Appa College – MS University.

MASTER DATA SCIENCE PROGRAM - GUVI

CAREER HISTORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Technology** | **Designation** | **Duration** |
| Colan Infotech Pvt. Ltd., | DataScience | Data Scientist | 2021 - Present |

PROJECTS

**Project : Hospital Staff’s Salary Prediction**

**Skills- Python, Pandas, EDA, Matplotlib, Seaborn, ScikitLearn**

* Developed a salary prediction model for hospital staff by Model development.
* Worked with a large dataset, performing data cleaning, transformation, and exploratory analysis.
* After testing different models, I chose **XGBoost Regressor** and gives high **accuracy 89%**.
* I worked to build **CI/CD** pipeline using **AWS** tools, which allowed for automated training, testing, and deployment of the model.
* Used S3 for data versioning, GitHub for code, SageMaker for training and deployment, Cloud Watch for monitoring, Step Functions for system testing, and CodePipeline for automation.

**Project : Insurance Fraud Prediction**

**Skills- Python, EDA, Matplotlib, ScikitLearn, AWS S3, DVC, Git, AWS SageMaker, MLFlow**

* I worked on detecting fraudulent insurance claims to help companies reduce financial losses.
* I Managed large dataset, addressing missing values and removing irrelevant columns.
* Finalized the **Random Forest** model for its **0.88 recall and 0.90 AUC-ROC**, effectively handling imbalanced data.
* Built a CI/CD pipeline with AWS tools to automate model training, deployment.

**Project : Custom AI Avatar Bot**

**Skills**: **Python, Transformers, Machine Learning, SQL, Xamp, Flask**

* Developed an intelligent chatbot with a speaking Avatar Interface to provide dynamic responses based on database queries.
* Built a Machine Learning model to classify user queries and determine the relevant database for information retrieval.
* Developed a **Transformer-based SQL query generator** to dynamically formulate accurate queries based on user questions.
* Implemented a **human-like response generator** that adapts to user input and retrieved data for natural, context-aware conversations.

**Project : LSG - Bert Text Classification**

**Skills- Python, SQL, BERT, Streamlit, Flask, Azure**

* Developed an Invoice Analyzing & Tracking Model using BERT-based text classification to categorize invoice descriptions and track legal compliance.
* Fine-tuned the **BERT model** to classify invoice descriptions into GNCCode.
* Implemented a tracking system to classify invoices as legal, illegal, or other service categories based on predicted GNCCode.
* Developed a custom model training architecture for user-preferred classes.
* Deployed training and inference apps on an **Azure Virtual Machine** and integrated **Azure Blob Storage**, resolving auto-restart issues with a service file.

**Project : Text Image Classification for OCR Project**

**Skills- Python, CNN, Tensorflow, Selenium**

* Developed a custom OCR application for extracting text from images.
* Segmented the problem into Handwritten Text vs. Machine-Printed Text classification.
* Designed and trained a Custom CNN model to classify text images into handwritten or machine-printed categories.
* Collected a diverse dataset using Selenium-based web scraping to enhance model.
* Experimented with various model architectures and hyperparameter combinations, incorporating callback techniques to optimize performance and prevent overfitting.

**Project : Resume Extraction System**

**Skills- Python, Spacy, BERT, Flask**

* Developed a resume parsing system integrating SpaCy and BERT to extract key entities.
* Applied NLP techniques for precise entity extraction and classification.
* Added NER features to extract PCI keywords for the network plots.
* Deployed the system using Flask, enabling seamless API integration with external applications.