HEMANTH KUMAR REDDY TIYYAGURA

+1 (703) 975 7183 | hemanth@mymailsbox.net | nanihemanth.github.io/Hemanth-Portfolio | linkedin.com/in/hk-tiyyagura

Highly motivated and analytical Data Scientist with 4 years of experience with a strong foundation in statistics, machine learning, and programming & AI.

EDUCATION

George Mason University, *MS in Computer Science - Machine Learning* | Virginia, USA GPA: 3.63 Dec 2023

SASTRA University, *B.Tech in Computer Science* | Tamilnadu, India GPA: 3.78 July 2020

EXPERIENCE

Capgemini , *Data Scientist* | Remote, USA 08/2023 - Present

- Spearheaded acquisition and preprocessing of 50+ datasets totaling 10 terabytes, employing Python and SQL to ensure data integrity and suitability.
- Produced 100+ charts and graphs using Matplotlib and Seaborn, facilitating clear presentation of insights gleaned from exploratory data analysis.
- Collaborated with senior data scientists to develop predictive models leveraging Regression, Decision Trees,and Random Forests, resulting in a 20% improvement in accuracy through feature engineering.
- Orchestrated model evaluation using cross-validation techniques, achieving average metrics of 90% accuracy, 85% precision, 88% recall, and 87% F1-score.
- Automated data preprocessing, model training, and prediction tasks through Machine Learning pipelines,enhancing operational efficiency by 30%.
- Regularly delivered actionable insights and trends to stakeholders via comprehensive reports, ensuring alignment with organizational objectives.

Cognizant Technology Solutions , *Programmer Analyst* | Chennai, India 08/2020 - 11/2021

- I spearheaded a transformative project that harnessed a suite of AWS services to build a robust and scalable RESTAPI for retrieving vehicles sales data and ensured seamless integration of AWS components for secure, reliable, and efficient retrieval of sales data.
- Designed and implemented end-to-end ETL pipelines to extract, transform, and load sales data from diverse sources using Informatica tool.
- Implemented CloudFormation templates to automate data processing workflows, resulting in a 10% improvement in productivity.
- Leveraged SQL expertise to investigate and diagnose data discrepancies, anomalies, and performance bottlenecks within the Tableau data sources.
- Employed AI-driven machine learning models to analyze historical sales data and generate accurate sales forecasts by tuning hyperparameters.
- Utilized advanced AI techniques, including deep learning and ensemble methods, to capture complex patterns and dynamics in sales data, resulting in a 30% improvement in forecasting accuracy and developed an interactive sales performance dashboard using seaborn,plotly,matplotlib

Novartis, *Data Engineer* | Hyderabad, India 07/2019 - 07/2020

- Led the design and implementation of end-to-end data pipelines leveraging Azure Data Factory, Azure Databricks, and Azure Synapse Analytics to streamline the ingestion, processing, and transformation of diverse pharmaceutical datasets.
- Implemented performance optimizations within Azure Databricks, resulting in 30% increase in data processing speed for critical analytics workflows.
- Utilized AI algorithms for feature extraction, dimensionality reduction, and feature selection, enabling the capture of nuanced patterns and relationships within the data. Designed and implemented predictive models using Python, focusing on classification and clustering algorithms.
- Implemented advanced AI algorithms, including Natural Language Processing (NLP) to uncover hidden patterns and trends within unstructured pharmaceutical data and techniques such as data normalization, feature scaling to handle missing values for ensuring data quality.
- Leveraged state-of-the-art Machine Learning Algorithms, including Deep Learning Architectures(RNN), to build predictive models capable of identifying potential adverse events with high accuracy

Tech Vedika INC, *Data Analyst* | Hyderabad, India 11/2018 - 06/2019

- Designed and created interactive Tableau dashboards to visualize Employee Attrition data trends, enabling Employers to make data-driven decisions.
- Utilized SQL queries to extract and transform Employee Attrition data from relational databases, ensuring data accuracy and completeness.
- Conducted extensive Data Preprocessing and Feature Engineering to extract relevant features from large-scale transactional datasets, enhancing the performance of the fraud detection models.
- Utilized advanced analytics and AI techniques, including Machine Learning algorithms such as Logistic regression, Random Forests, and Neural Networks, to detect anomalous patterns indicative of fraudulent activity.

SKILLS

Languages & IDE	Python, SQL, SAS, R, DART, VBA, Java ,Google Colab, Spyder, Pycharm, JupyterLab,Visual Studio Code
Data Visualization	Tableau ,Power BI, ArcGIS, Alteryx,Microsoft Excel,Powerpoint,SAP,Sharepoint,Google Sheets,A/B Testing
Tools	Orchestration(Docker, Kubernetes, Apache Airflow), ETL(Informatica, IICS),Streaming(Apache Kafka),Git
Azure Cloud	Databricks, ADF, Data Lakes, Azure Blob Storage, ADLS Gen2,Azure Synapse,Delta Lakes,Pyspark,Sparksql
AWS & GCP	EC2,S3,Redshift,Glue,EMR,Lambda,API Gateway,Dynamo DB,BiqQuery,Dataflow,Hadoop,HDFS,Spark
Machine Learning	Random Forest, Decision Tree,Classification,Clustering,Scikit Learn,Time Series, Numpy, Pandas,Django
Deep Learning	RNN,CNN,LSTM,Spacy,Keras,Pytorch,Tensorflow,NLTK,NLP,Flask,LLM,Vector DB,LlamaIndex,LangChain

PROJECTS AND PUBLICATIONS

Safe URL Inspector – *(Python, Machine Learning, Selenium, SSL, Node JS ,Tableau,Github)* 09/2023 – 11/2023

- Developed a comprehensive website and mobile app utilizing regular expressions to extract URLs from text messages and emails, Selenium for conversion of short-form URLs to their long-form equivalents
- Utilized a Random Forest classifier, incorporating features such as page rank, the number of question marks, and semicolons in URLs, to classify URLs as genuine or fake

LLM for Sentiment Classification - *(Python,NLP,BERT,Encoders,Decoders,Attention)* 07/2023 – 11/2023

- Implemented and fine-tuned a BERT model for sentiment classification of African tweets(12 different languages), achieving accuracy of 75% on African tweets .Successfully optimized hyperparameters for improved performance
- Contributed valuable insights into sentiment trends across African regions, demonstrating proficiency in NLP, Machine Learning model development, and data preprocessing.

Garbage Classification Website - *(Flask,Python,CNN,Tableau,Pytorch,HTML,Github)* 03/2023 – 04/2023

- Developed an innovative Flask-based web application that utilizes CNN to classify waste materials into recyclable or organic categories with PyTorch to achieve accurate waste classification(92% accuracy)
- The user-friendly web interface allows individuals to upload images of waste items and receive instant predictions, promoting sustainable waste management practices and raising awareness about recycling.Integrated Tableau for data visualization, presenting waste management insights and trends

Facial Emotion Recognition – *(Python, CNN, Scikit, Pytorch, Computer Vision)* 01/2021 – 07/2021

- Published a research project in Springer, presenting a comprehensive facial recognition system that integrated Haar Cascade for facial detection, data augmentation to enhance dataset diversity, and CNN for emotion classification
- The proposed model detects human face in image or video and Emotion Classification is performed with 80% accuracy achieved through CNN.IEEE Publication Link

CERTIFICATES & ACHEIVEMENTS

- Achieved recognition in Blue Clarity "Bring Down Counterfeiting-2023" hackathon, winning the prestigious Best Overall Solution Student Prize and earning the Crowd Source Prize **Hackathon**
- Microsoft Certified: **Power BI Data Analyst Associate**
- Microsoft Certified: **Azure Data Engineer Associate**