Hemanth Kumar Reddy Tiyyagura Data Scientist

hemanth@mymailsbox.net (703) 975-7183 Charlotte , NC Linkedin Portfolio

Profile

Highly motivated and analytical Data Scientist with 3+ years of experience with a strong foundation in statistics, machine learning, and programming. Equipped with hands-on experience in data analysis, data visualization, and model development using Python and SQL. A quick learner and team player, committed to applying data-driven insights to solve complex problems and make informed business decisions.

Skills

Languages and IDE (Python, SQL, HTML, CSS, C, VBA, Google Colab, Spyder, Pycharm, Jupyter Lab, Visual Studio Code, SAS, R)

Cloud & Big Data (Microsoft Azure (Databricks, ADF, Data Lakes, Delta Lakes, Azure Blob Storage, ADLS Gen2,Azure Synapse), AWS (EC2, S3, Redshift, Glue, EMR, Lambda, API Gateway, Dynamo DB), GCP(Big Query, Dataproc), Snowflake, Hadoop, HDFS, Apache Spark, Apache Airflow, PySpark, Spark SQL, Map Reduce, Apache Kafka, NoSQL(Database))

Tools (Tableau ,Power BI, ArcGIS, Alteryx, Docker, Kubernetes, Informatica, IICS, Microsoft Excel, PowerPoint, SAP)

Frameworks (Django, Streamlit, Flask, Keras, Pytorch, Tensorflow, NLTK, GenAI (Vector DB, Llama Index, Lang Chain, Vertex AI),PostgreSQL, SQL Server)

Miscellaneous (Spacy, Scikit Learn, Time Series, Numpy, Pandas, Statistics (A/B Testing, Probability) ,Statistical Analysis, Parquet, Jira, SSIS, DevOps)

Professional Experience

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

- 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.

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

- 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.
- Integrated AWS services including API Gateway, CloudFormation, Lambda, and RDS for efficient customer data retrieval, IAM for user access control.
- Leveraged SQL expertise to investigate and diagnose data discrepancies, anomalies, and performance bottlenecks within the Tableau data sources.
- Utilized Amazon EMR (Elastic MapReduce) for distributed data processing, employing PySpark and Python to perform complex transformations and aggregations on large customers sales dataset and developed an interactive sales performance dashboard using seaborn, Plotly, Matplotlib.

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

- 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.
- Designed and developed interactive dashboards and visualizations in Tableau and Power BI to uncover trends and patterns in pharmaceutical sales, production, and clinical trial data. Utilized line charts, bar graphs, and heatmaps to visualize sales data and identify actionable insights for sales team
- Implemented performance optimizations within Azure Databricks, resulting in 30% increase in data processing speed for critical analytics workflows.
- Developed and trained Machine Learning models using Python libraries such as scikit-learn and pytorch to analyze pharmaceutical data and extract actionable insights. Implemented techniques such as data normalization, feature scaling, and handling missing values to ensure data quality
- Implemented data governance policies and security measures to protect sensitive pharmaceutical data and ensure compliance with regulatory requirements. Established clear guidelines and procedures for data collection, storage, access and disposal to ensure accountability and transparency.

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

- 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 techniques, including Machine Learning algorithms such as Logistic regression, Random Forests, and Neural Networks, to detect anomalous patterns indicative of fraudulent activity.

Education

Master of Science, *George Mason University-Fairfax* 01/2022 – 12/2023 | Virginia, USA
Computer Science(ML Concentration)

Bachelor of Technology, *SASTRA University* 07/2016 – 05/2020 | Thanjavur, India
Computer Science

Projects

Safe URL Inspector, *Python, Machine Learning, Selenium, SSL, Node JS ,Telegram Bot, Tableau* 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.

Facial Emotion Recognition using Haar Cascade and CNN, 01/2021 – 07/2021
Python, CNN, Scikit, Pytorch, Computer Vision (IEEE Publication)

- 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. IEEE Publication Link: (Tap)
- The proposed model detects human face in image or video and Emotion Classification is performed with 80% accuracy achieved through CNN.

Achieved recognition in Blue Clarity
"Bring Down Counterfeiting-2023" hackathon, winning the prestigious Best Overall Solution Student Prize and earning the Crowd Source Prize

Certificates

- Microsoft Certified: Power BI Data Analyst Associate
- Microsoft Certified: Azure Data Engineer Associate