AnilkumarReddy Chittem

240-4096962 | anilkumarreddy.chittem@gmail.com |https://www.linkedin.com/in/anilkumarreddychittem/ | https://github.com/Anilkumar-Chittem

Summary

- Experienced **Data Science and MLOps** professional with 6 years of experience with a strong mathematics, statistics, and programming background. Proficient in programming languages such as Python, MATLAB along with tools like Simulink, PowerBI, Tableau, and Azure for data analysis and visualisation.
- Skilled in data engineering, data cleaning, preprocessing, exploratory data analysis, designing and managing scalable infrastructure, deploying, and serving models, building robust data pipelines, implementing CI/CD processes, and applying statistical techniques and machine learning algorithms.
- Effective communicator with a track record of delivering actionable insights and driving data-driven decisions, as evidenced by conference publications and impactful contributions to projects.

- Skilled in applying statistical techniques and machine learning algorithms to develop models that can predict outcomes, classify data, segment populations, and make data-driven decisions.
- Experience applying a wide range of machine learning algorithms and developing various **Regression, Classification, Time series, Anomaly detection, Propensity, Market Mix Models, Multi-touch attribution, NLP, Deep Learning, and LLM models.**
- Experienced in developing and implementing machine learning algorithms (supervised and unsupervised learning), including hierarchical clustering, KNN, SVM, LSTM, CNN, and more, to optimize network performance and enhance customer churn forecasting.
- Skilled in utilizing advanced tools and frameworks such as Hugging Face, LLMs, PyTorch, and TensorFlow for natural language processing and chatbot development.
- Proficient in data visualization using Power BI, Tableau, and Qlik Sense, with a proven track record of designing interactive reports that showcase KPIs, trends, and anomalies.
- Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in Azure Data bricks.
- Developed Spark/Scala, Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources.
- Data sources are extracted, transformed, and loaded to generate CSV data files with Python programming and SQL queries.

- Proven ability to lead research projects and collaborate with multidisciplinary teams to develop innovative solutions, as shown in projects like predicting cybersecurity threats and fault conditions in nuclear power plants.
- Effective communicator with a track record of delivering actionable insights and driving data-driven decisions, as evidenced by conference publications and impactful contributions to projects.

TECHNICAL SKILLS

Project Management tools	Microsoft Project, Jira, Confluence, Agile, Scrum, SharePoint
Database	MS SQL, Big Query, Postgre SQL
Programming	SQL, T-SQL, Python (Pandas, NumPy), R, PySpark, PySQL, SnowSQL
Versioning Tools	Git, Docker, DVC
MLOps tools	MLFlow, AutoML, Azure MLOps, Azure ML Designer
Machine Learning Algorithms	Classification, Regression, Forecasting, Decision Trees, K-Means, SVM, Boosting and Bagging methods, **NLP (Sentiment Analysis, Semantic search)**, **Market Mix Modelling (MMM),** Multi-touch attribution models, Generative AI, and Deep Learning. LLM (BERT/GPT/T5), Prompt engineering
Machine Learning Libraries	NumPy, Pandas, PyCaret, Scikit-learn, TensorFlow, PyTorch, Keras, Transformers, SpaCy, Matplotlib, Seaborn
Statistical Methods	A/B Testing, T-test, Chi-square test Predictive Analysis, Hypothesis Testing, PCA, MBA, Text Analytics, ANOVA, Text Mining (Bag of Words, TF-IDF)
Data Visualization	Power-BI, Tableau, Google Looker Studio
Cloud Computing	**Azure (DevOps, Data Lake Storage, ML Studio, Data Factory, Data Bricks, Synapse), Amazon Web Services (EC2, S3, Glue, Lambda, Sage Maker, EMR), GCP (ML, BigQuery,), Snowflake**

PROFESSIONAL EXPERIENCE

App Orchid – Data Science Intern, *USA* November 2023 – May 2024

- Analysed specialised text to SQL datasets, implementing precise evaluation metrics that increased query accuracy trend by 25% and reduced data retrieval time by 15%.
- Learned a prompt engineering technique called Chain-of-Thought on a product feature that utilises LLMs to convert text into structured pipelines in JSON format and got a task-specific accuracy of 87%.
- Engineered robust Selenium web scraping solution to systematically harvest procurement contract lists from key websites, resulting in a 50% reduction in manual labour and a 20% increase in data comprehensiveness.

Machine Learning for Customer Churn Forecasting.

- Project Description: Made data-driven decisions in implementing advanced machine learning models, including Random Forest, Gradient Boosting and Neural Networks for customer churn forecasting. Achieved a precision of 85% and a recall of 80%, leading to a 20% reduction in churn rate.
- Responsibilities:
- Developed and implemented machine learning models, including Random Forest, Gradient Boosting and Neural Networks for customer churn prediction.
- Conducted data analysis and feature engineering for model development.
- Utilized SQL, Python, and TensorFlow on Azure AI Platform for advanced analytics and model.

Automated PDF Extraction for Digital Twin Development.

- Project Description: Automated PDF extraction from Structural Analysis documents using PyMuPDF and Alteryx generate 3D tower representations in Revit using Dynamo Script, contributing to Digital Twin development.
- Responsibilities:
- Utilized Alteryx workflows to preprocess PDF documents and extract structured data.
- Integrated Alteryx with PyMuPDF for PDF parsing and data extraction.
- Developed scripts in Dynamo Script to convert extracted data into 3D representations in Revit.
- Utilized Python, Alteryx, PyMuPDF, and Dynamo Script.

Hierarchical Clustering for Network Optimization.

- Project Description: Applied hierarchical clustering alongside advanced anomaly detection algorithms including KNN and SVM to optimize network performance, resulting in a 25% improvement in network efficiency and a 15% reduction in network downtime.
- Responsibilities:
- Implemented hierarchical clustering algorithms for network optimization.
- Utilized KNN and SVM for anomaly detection and network performance improvement.

Yupptv - Data Science Associate, *Hyderabad* August 2021- July 2022

- Collaborated with a Group of 3 Full Stack Developers and a Junior Data Scientist to execute cutting-edge adaptive testing algorithms on a newly designed entrance exam portal using Python, Flask, SQL, HTML5, CSS3; boosted client business engagement capabilities by 30% efficiency and forecasting 90% testing accuracy.
- Leveraged deep domain knowledge and expertise in data science, MLOps, and data engineering on the Azure Cloud platform to provide strategic guidance and insights on complex data-driven projects.
- Collaborated with cross-functional teams to develop a GenAI HR application for Job profile creations leveraging OpenAI models such as GPT-4 and LangChain as a framework to generate highly contextual and domain-specific content.
- Developed and implemented a propensity model framework using XGBoost on Azure Databricks to predict the likelihood of conversion enabling business to prioritize leads that increased sales revenue by 26% through accurate forecasting of market trends.
- Worked on building customer segmentation models using K-means clustering and DBSCAN machine learning techniques utilizing Adobe Analytics through Python helping businesses to identify the target audience and tailor personalization campaigns.
- Worked on the development and deployment of a demand forecasting model using deep learning techniques, Employed LSTM neural networks within TensorFlow and Keras to model and predict future product demands based on historical sales data,
- Partnered with the Marketing Analytics team from AdTech and retail clients in the US to build a Market Mix Model (MMM) and multi-touch attribution model using Lightweight MMM using Azure ML Studio to quantitatively estimate the effectiveness of various marketing elements, identify key sales drivers and the high ROI Channels.
- Utilized Apache Hive and Spark to manage and analyze large datasets in a predictive maintenance machine learning with a focus on developing and optimizing ETL processes in Hive to structure and query big data efficiently, while employing Spark's MLlib in Databricks for building and training predictive models.
- Developed and managed a comprehensive data pipeline in Azure Databricks, integrating diverse source systems to

populate a feature store, enabling efficient model development and deployment.
- Worked on deployment and optimization of a Convolutional Neural Network (CNN) model using TensorFlow for product image classification task, leveraging Azure cloud platform capabilities. This initiative streamlined the automation of product categorization, significantly improving operational efficiency and accuracy in inventory management.
- Used Spark 2.0 (PySpark) to develop a variety of models and algorithms for analytic purposes.
- Performed and treated outliers and missing values detected using boxplots and Pandas predefined functions.
- Worked with dimensionality reduction techniques like PCA, LDA, and ICA.
- Worked on Teradata SQL queries, Teradata Indexes, and Utilities such as Multi load, Tpump, Fast load, and Fast Export.
- Configured Hadoop tools like Hive, Pig, Zookeeper, Flume, Impala, and Sqoop.
- Used Data Warehousing Concepts like Ralph Kimball Methodology, Bill Inmon Methodology, OLAP, OLTP, Star Schema, Snow Flake Schema, Fact Table, and Dimension Table.
- Refined time-series data and validated mathematical models using analytical tools like R and SPSS to reduce forecasting errors.

Flip Robo Technologies – Data Science Intern, *India (Remote)* August 2017 – July 2021

Predicted The Crime Rate Of Baltimore City Based On Features Of The Day.
- Advanced a predictive analysis project on Crime volume in a Baltimore City police organisation, leveraging 558,347 crime data points and 58,476 weather observations; enhanced crime prediction accuracy by 25%.
- Management with 2 interns to functions 8 regression models; LightGBM initiated the highest score of 0.49. Day features and weather show a weak uncertainty with crime volume.

Semantic Segmentation Of Indian Traffic Dataset Using Convolutional Neural Network.
- Led the supervision of semantic segmentation for Indian traffic using a 5.5 GB dataset in which a computer vision task involved images captured by a consumer front-facing car camera.
- Innovated with complex image augmentation techniques, producing 25,000+ diverse training examples for Data analysis that preserved semantic clear content and increased model precision by 22% .
- Engineered a U-Net with VGG19 for semantic segmentation, modifying the encoder and achieving convergence in 42 epochs with 29,060,982 trainable clear parameters; improved IoU scores.

Abuse Comment Classifier.
- Deployment a clear application for an online hate and abuse comment classifier; identified 95% of offensive content and reduced instances of online risk harassment feedback by 40%.
- Governance an initiative to evaluated abusive language through EDA, stop words removal, and word clouds; Established Decision Tree Classifier model, resulting in a 95% test accuracy output pattern and a 92% F1 score.

Global Regulatory Information Portal.

- Led an Agile team of 5 members offshore including Developers, Testers, Database Analysts, and served as the single point of contact for the project from onsite.
- Gathered requirements, created user stories and wrote acceptance criteria. Maintained sprints and backlog using JIRA.
- Developed models using decision trees, random forest, logistic regression.
- Analysed data and created dashboards, reports and performed data visualization using Tableau.
- Achieved 15% increase in productivity by analysing the performance data of the team.

Data Acquisition Framework and Predictive Cybersecurity for Edge Devices.
- Led the design and development of the data acquisition framework, ensuring compatibility with edge devices.
- Integrated Electronic Control Modules with CANBus and SCADA for seamless data exchange and control.
- Implemented predictive cybersecurity measures to protect edge devices from potential threats.
- Directed the modelling and simulation of nuclear power plants using Logistic Regression, LSTM, CNN, and Exponential degradation models, leading to improved fault detection and diagnosis, predicting future failures and remaining useful life.
- Achieved a reduction of annual inspection and maintenance labor costs by 50-75%.

EDUCATION

University of Maryland Baltimore County, MD
Master of Professional Studies: Data Science GPA- 3.9

RESEARCH EXPERIENCE

- Enhance mobile photography to audit blur images in low lighting conditions, with 500 samples from HDR and Exdark datasets mainly curiosity on Gaussian blur.
- Conducted GANs with PCA modelling on the Exdark dataset, resulting in 15.29 PSNR and 0.285 SSIM; merging datasets with CNN yielded findings 13.25 PSNR and 0.412 SSIM, enhancing overall image quality metrics.