BOMMIREDDY VIJAY KUMAR REDDY

+1 (317) 970-6752 | vikbommi@iu.edu | LinkedIn | Tableau

EDUCATION:

Indiana University-Purdue University | Indianapolis, IN — Exp Grad – Dec2024

Master of Science in Applied Data science — GPA: 4.0/4.0

Coursework –*Scientific Clinical data mgmt, Intro to informatics, Intro to Bio stat-R, data visualization, cloud computing and math.*

SKILLS:

Programming & Tools : Python, R, SQL, Gitlab, Git, HTML , Power BI, Microsoft PowerPoint.

Libraries : NumPy, Pandas, Matplotlib, Seaborn ,PySpark, Genism, Beautiful soup, PyTorch, Stats model, Spacy, ggplot2.

Data bases &Cloud : MySQL, Microsoft SQL Server, MySQL, NoSQL, MS Access &AWS(Redshift, Athena, Glue, S3).

Analytical Techniques : Data Preprocessing, Feature Engineering, Inferential Statistics, Modeling, Ensemble Methods, ETL Pipelining, Classification, regression, clustering, Data Mining, Neural Networks.

WORK EXPERIENCE:

COMET Lab at IUPUI | Indianapolis, IN — Jan 2023-Present

Research Assistant

- Utilized Pandas to perform in-depth analysis on claims data for **1M patients** and translated findings into interactive Tableau dashboards.
- Developed predictive models using scikit-learn to estimate the impact of proposed preventative care incentives and conducted data and fine-tuning model parameters validation to address potential overfitting.
- Modeling and regression analysis demonstrated a potential **$1.5M** in savings through expanded employer-sponsored wellness initiatives.

Innovation, Incubation & Entrepreneurship Development Centre - NIT Srinagar| India — Dec 2021-Dec 2022

Data scientist

- Executed data cleaning and **statistical testing** on large dataset, made an interactive Tableau dashboard to showcase results.
- Wrote a Python script to accurately analyze research study data by handling repeating instruments, categorizing various family groups, and optimizing computational efficiency.
- Collaborated with **6+ cross-functional teams** to design and deploy dashboard, improved operational efficiency and data quality.
- Through deep learning models in conjunction with Fuzzy logic, employing Keras, TensorFlow, and OpenCV libraries, achieved an accuracy exceeding 90% across in **[ICESC-195]**, **[ICIRCA-400]**, **[ICIRCA-282]**, all three research publications.

Coincet.ai |India — Jul 2021-Dec 2021

Data science intern

- Using SQL and Excel **pivot tables**, interpreted large sales datasets to extract actionable insights tailored to business needs.
- Implemented data cleaning, feature engineering, and exploratory analysis in Python on an **8M+ row** database to enable modeling.
- Analyzing requirements, adeptly generated tailored ad-hoc reports using Excel showcasing agility in delivering precise insights.

ACADEMIC PROJECTS:

Linguistic-Based Intelligent Phishing Email Detection| *Data Wrangling*| *Natural Language Processing.*

- Wrangled meta data and analyzed email content to engineer multidimensional features through Named Entity Recognition.
- Established an intelligent dynamic link analysis system which blacklists malicious links without relying on phishing links database.
- Applied XGBoost based classification model with hyper-parameter tuning using **10-Fold Cross Validation** and achieved the best AUC- ROC score of **0.95 on test data**.

Tesla Stock Analysis and Prediction |*Python*| *ML in Finance*| *Excel* |*Regression Analysis*.

- Examined effects and gathered data for macro-economic factors, competitors, and exchange rates affecting stock prices.
- Build financial models using statistical methods, autoregressive, linear, and decision-based algorithms with **98% accuracy**.
- Gained **18% profit** by generating buy & sell trading signals using simple and exponential Moving Average Crossover strategy.

Home Credit Default Risk |*Advance excel*| *Machine learning.*

- Integrated seven datasets comprising over **20 million rows and 200+ columns**, conducting extensive feature engineering and exploratory data analysis to meticulously prepare the data for modeling.
- Formed a comprehensive Machine Learning pipeline to construct a classification model for predicting Credit Default Risk, employing ensemble learning methods.
- Provided interactive visualization approaches, utilizing tools such as Matplotlib and Seaborn, to give informative representations of data patterns and model predictions, improving stakeholders' interpretability.