Bhanu Prakash Reddy

bhanuprakash@jobemails.net (214) 218-6957 Dallas, Texas LinkedIn

SUMMARY: Dynamic and results-oriented Data Scientist with over 5+ years of experience in leveraging data-driven insights to drive business improvement and innovation. Proficient in a range of advanced analytical techniques and tools, including machine learning, statistical analysis, and big data platforms. Demonstrated expertise in developing and implementing predictive models and algorithms that have successfully informed business strategies and solutions.

TECHNICAL SKILLS

Languages & Databases	Python, R, SQL, MS Excel, spark, flask, Docker
Visualization Tools & Others	Tableau, Power BI, MS PowerPoint, NoSQL, MongoDB, hive
Python and R Libraries	Pandas, NumPy, Scikit-Learn, NLTK, Genism, TensorFlow, PyTorch, Matplotlib, Seaborn, Plotly
Machine Learning	Statistical modeling, Linear Regression, Logistic Regression, Naive Bayes, Lasso & Ridge Regression, Linear/Quadratic Discriminant Analysis, Decision Tree, Random Forest, XGBoost, SVM, K-Means, Word2Vec, BERT, CNN, RNN, LSTM, Bootstrap Sampling, Cross-Validation, Hypothesis Testing, PCA, SVD, Collaborative Filtering, A/B Testing, Clustering, NLP, Text Mining, ARIMA, SARIMA, Association Chain Mining.
Certifications and Courses	**Google Cloud Professional Machine Learning Certification**

PROFESSIONAL WORK EXPERIENCE:

BNY Mellon, Remote, USA **June 2023 – present**

Data Scientist

- Created **ETL data mining pipelines**, processing more than 1M records/day, **optimized data storage** & **reduced previous run-times** by 30% through **compression techniques** (Zstandard), **indexing** (B-trees, bitmap indexes).
- Developed **real-time data pipelines utilizing Apache Kafka** to feed data into ML models and generate predictions in near real time.
- Developed **Random Forest Regressor, VARMAX, LSTM, Prophet, and KNN time series models** on 150,000 data with feature engineering to forecast End of Day Balances and File Row Count **with R2 scores of 0.9983 and 0.9938** respectively for best performing **Random Forest Regressor model** using Python, MLOps, and GitLab.
- Built the charge off model – used **LightGBM**, to improve the collection strategies. Model predicts the likelihood of an account going to Charge off state in next 12 months.
- **Scheduled jobs in Airflow and SQL** to regularly generate daily insights, data backhauling and data quality checks.
- Designed **Power BI dashboards** to analyze data & derive business insights by **reporting quantitative KPIs** and enabling self-service **root cause analysis**; reduced fail rate by 38% and improved operations efficiency & client experience.

Yoamigos Webservices, India **Jul 2017 – Jul 2022**

Senior Data Scientist

- Refined and augmented an **XGBoost model to forecast patient disease**, achieving an 8% improvement in prediction accuracy by feature engineering, time frame adjustments, and leveraging **Explainable AI (XAI)** to enhance the model's interpretability.
- Forecasted sales for more than **600 products across 5 retailers** in multiple countries resulting in trade promotion optimization (TPO). The built **Auto Regressive Distributed Lag (ARDL) model** achieved **13% WMAPE** that far exceeded the accuracy of manual forecasts.
- **Customer segmentation using RFMT model:** Applied the Recency, Frequency, Monetary, and Time (RFMT) model to analyze patient data. Employed **K-Means, Gaussian, and DBSCAN algorithms** to classify patients into distinct groups. Conducted cluster factor analysis using methods such as **elbow, dendrogram, silhouette, Calinsky–**

Harabasz, Davies–Bouldin, and **Dunn index** to ensure the meaningfulness of patient groupings. Implemented the **majority voting (mode version) technique** to select the most relevant patient clusters.

- Built **BIRCH clustering-based anomaly detection tool** for early detection of file count explosions on Data Lake.

Data Scientist

- **Designed a payment integrity solution** to classify claims into different levels of risk using **Support Vector Classifier (SVC)** with a **recall of 91% at AUC of 0.95.**
- **TV Rating Estimation**: Identified key factors those impact client's TV ratings and estimated these ratings with an **Adj-R2 ≈ 0.92 & MAPE ≈ 4.5%**; thereby aiding the client to sign a $130 million deal with broadcasters.
- **Digital Transformation using Google Tag Manager:** I seamlessly incorporated and managed over **30 tracking tags**, resulting in a **remarkable 40% boost in tracking efficiency.** Additionally, I significantly improved conversion tracking precision, leading to a **25% increase in appointment bookings**, a **30% rise in newsletter sign-ups**, and a **20% surge in resource downloads**.
- **Acquisition Campaign Planning**: Developed a framework and measured the effectiveness of marketing campaign run on Radio; Deployed **A/B testing** to compare pre- vs post-campaign cost per order; Identified best Metropolitan Statistical Areas (MSA) responsive to campaign and **saved $200K future marketing spends**.
- **Strategic Market Research and Data-Driven Marketing Analysis:** Employing rigorous **cross-referencing techniques**, I uncovered latent market insights from NAICS, USA.gov, and U.S. Census data. Using **statistical analysis, regression modeling, and sentiment analysis**, I identified hidden patterns in industry trends and patient preferences. By seamlessly integrating consumer statistics, demographics, and economic indicators, I precisely tailored our initiatives to match customer demands, resulting in a remarkable 25% growth in market share.

ACADEMIC PROJECT EXPERIENCE

- Utilized **BERT model to identify SEO keywords** and built a model to rate their importance based on search volume and **rank using cosine similarity**. **Conducted prompt-engineering experiments in GPT3/GPT4** based models from OpenAI to minimize hallucinations and improve text generation.
- **Bankruptcy Prediction (5th among 27 teams):** Built classification model utilizing **Gradient Boosting** to predict financial distress. Employed SMOTE to over-sample unbalanced classes and achieved ~ 0.93 AUC.
- **Content Moderation**: Detected hate speech to moderate content on Craigslist discussion forums. Gathered unstructured data by **leveraging web scraper** and **applied sentiment analysis** to achieve ~ 0.76 AUC.
- **Text Summarization**: Engineered a deep learning-based model utilizing advanced NLP techniques, including **LSTMs**, to generate concise and coherent summaries of long articles with a ROUGE-2 score of approximately 0.73.
- **Keyword Extraction**: Created an automated system that extracts the most important keywords from large bodies of text using advanced NLP techniques and machine learning algorithms, resulting in a high F1 score of approximately 0.85.
- **Database Performance Tuning**: Optimized the performance of a database system for a healthcare provider, reducing query response time by 60% and increasing system throughput by 40%. Utilized **indexing, query optimization, and database tuning techniques** to achieve performance improvements.

VOLUNTEER EXPERIENCE

National Service Scheme, President — Aug 2015 – Jul 2016

EDUCATION

- **Master of Science, Business Analytics (Data Science and Machine Learning), GPA 3.90**
  The University of Texas at Dallas — **Dec 2023**
  Relevant Courses: Applied Machine Learning, Modelling for Business Analytics, Time Series and Econometric Analysis, Data Visualization, Database foundations for Data Science, Big Data, Machine Learning in Google Cloud, Advance Statistics, Predictive Analytics and Prescriptive Analytics.

- **Bachelor of Engineering, BITS Pilani, India** — **Aug 2017**