

Bindu Latha Baniseti

baniseti.b@northeastern.edu | 857-867-2699 | [linkedIn/bindu-latha-baniseti](https://www.linkedin.com/in/bindu-latha-baniseti) | bindubanisetish@gmail.com

EDUCATION

MASTER OF SCIENCE IN DATA SCIENCE

Boston, MA | Dec 2023

NORTHEASTERN UNIVERSITY

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE ENGINEERING

Vijayawada, AP | Mar 2013

V R SIDDHARTHA ENGINEERING COLLEGE

Artificial Intelligence skills: Machine Learning, Deep Learning, Natural Language Processing, LLMs, Computer Vision, Probability, and Statistics

Programming skills: Python, SQL, Numpy, Pandas, Matplotlib, Tableau, Scikit-learn, PyTorch, R, AWS, GCP, PySpark, Docker, and Git

WORK EXPERIENCE

DATA SCIENTIST INTERN | SPARE-IT

Boston, MA | Jan 2023 – Aug 2023

INSTANCE SEGMENTATION USING COMPUTER VISION AND REPORT GENERATION USING GPT MODELS FOR SUSTAINABLE WASTE MANAGEMENT

- **Object Detection and Segmentation using YOLO, U-NET and Mask R-CNN:** Performed Instance segmentation by fine-tuning pre-trained models like Circular-Net on waste management dataset. Developed Lambda functions using AWS CDK to facilitate data pipelines.
- **LLM based summary Report generation from segmented images:** Finetuned GPT-based foundation model for abstractive summarization by employing prompt engineering, and Langchain leveraging past statistical reports, and customer queries. Increased overall customer satisfaction by 40%.

DATA SCIENTIST | DATAMETICA

Pune, MH | Aug 2018 - Dec 2021

AUTOMATIC TICKET CLASSIFICATION USING MACHINE LEARNING AND DEEP LEARNING

- **Word embeddings:** Cleaned customer texts using Stemming, Lemmatization, tag removal and other methods, and created word vectors using TF-IDF, word2vec, FastText and Glove to create features for Supervised classification.
- **Machine Learning models:** Utilized Naïve Bayes, Logistic Regression, CatBoost, and XGBoost models to classify customer tickets into 5 levels of severity and urgency. Added Human in the loop to handle exceptions and improved accuracy to 85%.
- **Deep Learning models:** Implemented RNN and LSTM models for features with more than a year of data. Applied techniques like Regularization, Early stopping, and Skip connections and achieved an accuracy of 91% AUC.
- **Dashboards using Tableau:** Designed real-time dashboards to monitor trends and anomalies and track metrics and KPIs and communicated insights and recommendations to stakeholders.

TEXT SUMMARIZATION USING CLASSICAL ML, DEEP LEARNING, AND TRANSFORMER MODELS LEVERAGING EXTRACTIVE AND ABSTRACTIVE STRATEGIES

- **Automated Summaries:** Developed Text summaries to help Banking clients prepare legal briefs concerning their legal positions. Created both Extractive and Abstractive summaries using various Deep Learning algorithms including BERT embeddings, Attention and Transformer architectures.
- **Extractive Summarization:** Used DL models like RNN, and LSTM for generating concise takeaways. Fine-tuned the models using pre-trained Transformer based BERT models for better accuracy among popular domains.
- **Abstractive Summarization:** Optimized LSTM model for generating accurate summaries by experimenting with various architectures and utilized pre-trained GPT model for fine-tuning.
- **Metrics:** Used metrics like Summary Length, Compression Ratio, ROGUE, and BERT Scores for optimizing across various domains and achieved an average accuracy of 89%

PROACTIVE CHURN PREDICTION: PREDICTIVE ANALYTICS FOR CHURN REDUCTION

- **Data wrangling and Feature Engineering:** Created data pre-processing pipelines to perform cleaning, outlier handling, normalization, and standardization, for customer features related to demographics, device type, plan type, billing methods and usage patterns.
- **Supervised Classification models:** Implemented Logistic Regression, Random Forest and XGBoost to classify potential churn profiles. Enhanced the model performance through Hyper-parameter tuning and Cross validation to achieve an F1 score of 93%.

DATA SCIENTIST | CAPGEMINI

Hyderabad, TG | Oct 2016 – Jul 2018

PREDICTIVE ANALYTICS FOR AUGMENTING CUSTOMER EXPERIENCE FOR BANKING PRODUCT (FINACLE)

- **Anomaly Detection to reduce Financial Fraud:** Developed an outlier detection model to flag fraudulent and circular transactions using transaction details, user history, user profiles and other features. Overcame the drawback of cold start problems using K-means and enhanced the model using Isolation Forest to achieve an accuracy of 92%.
- **A/B Testing to enhance consumer experience:** Conducted various A/B tests to address complaints regarding Automated customer support, Dashboard design and response, security protocols etc. to ensure proper alignment of product with expectations and ad-hoc improvements. Improved the KPIs for engagement and satisfaction score from 65 to 90%.

DATA ENGINEER | INFOSYS

Hyderabad, TG | Mar 2014 – Sep 2016

OPTIMIZING DATA INTEGRATION FOR REAL-TIME ANALYTICS AND REPORTING

- **ETL Pipelines for Reports and Analytics:** Developed and fine-tuned ETL script to speed up data ingestion into data warehouses to support real-time and batch data processing for dynamic dashboards.
- **SQL Query optimization:** Optimized queries using techniques such as Indexing, caching, column pruning, sharding, and locking for swift data retrieval and reducing server loads. Improved database performance by 24%.

CERTIFICATIONS & AWARDS

- **Machine Learning Specialization** and **Deep Learning Specialization** certification from **Deeplearning.ai**.
- **Google Certified Professional Data Engineer**.
- **PySpark** certification from **Udemy**.
- Endowed with annual **StarAward** in **Datametica** for successful deployments following CI/CD pipelines.