Bindu Latha Banisetti
banisetti.b@northeastern.edu | Boston, MA | 857-867-2699 | linkedIn/bindu-latha-banisetti | bindubanisettish@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.
- Dataset preparation for LLM fine-tuning: Extracted image coordinates and object information from output COCO files obtained from Circularnet and converted into descriptive texts. Manually created summaries for these texts with predefined guidelines providing a hierarchy of details highlighting the contamination rate and diversion rate.
- GPT2 based Abstract Summary through Auto-Regressive Text Generation: Fine-tuned pretrained GPT-2 using only Masked Multi Head Attention based Decoder as an Autoregressive next word predictor to create concise summaries. Trained with domain specific, diverse datasets to simplify complex technical jargon to increase downstream consumption for larger audiences.

DATA SCIENTIST | Datametica — Pune, MH | Aug 2018 - Dec 2021
Text Summarization using Attention-based Transformer models like BERT and T5

- Legal Summaries from Litigation documents: Leveraged Encoder and Decoder based Transformer models optimized with Attention mechanism to fine-tune pre-trained models like BERTSUM, Text-to-Text (T5), and GPT models to create automated Legal summaries.
- BERT based BERTSUM for Extractive Summary for Interpretability: Used WordPiece Tokenizer to create Token, Segment, and Positional embeddings to aid in capturing Semantic, Syntactic, and Morphological relationships between words to achieve deep and dynamic vector representations.
- Text-to-Text Transformer (T5) as Denoising Auto-Encoder for Abstractive Summary: Created SentencePiece based embeddings and customized Encoder and Decoder Layers with varying Masked Multi-Head Self-Attention and Multi-Head Cross-Attention blocks adjusting hidden dimension units connecting to Feed forward neural network.
- Metrics and Interpretability: Balanced trade-off between interpretability and accuracy with metrics like Summary Length, Compression Ratio, ROUGE, BLEU, scores for optimizing across various domains and achieved an average accuracy of 91%.

Automated 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 88%.
- 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 93% AUC.
- Dashboards using Tableau: Designed real-time dashboards to monitor trends, anomalies, track metrics, KPIs and communicated insights, actionable recommendations to stakeholders.

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 cold start problems using K-means and with Isolation Forest improved accuracy upto 92%.
- Proactive Churn Prediction using Supervised Classification: Developed 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%.
- A/B Testing to enhance consumer experience: Conducted various A/B tests to address complaints regarding Automated customer support, Dashboard design and 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 scripts 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 and PySpark certification from Udemy.
- Endowed with annual StarAward in Datametica for successful deployments following CI/CD pipelines.