Amey Basangoudar

(857) 250-1842 | basangoudar.amey@gmail.com| LinkedIn | GitHub

SUMMARY

Data scientist with hands-on experience in building scalable data solutions using advanced machine learning algorithms, statistical modeling, and data engineering techniques. Demonstrated ability to quickly grasp key performance indicators (KPIs) across diverse sectors, efficiently developing data-driven solutions for complex business challenges.

EDUCATION

Master of Science in Data Science - *Northeastern University (GPA 3.7), Boston, MA* ***Dec 2023***

Bachelor of Engineering in Electronics and Communication - *KLE Tech University (GPA 4.0), India* ***May 2020***

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, C++
Concepts: Statistics, Data Analytics, ML, AI, NLP, A/B Testing, Data Mining, Predictive Analytics, Data Visualization, Time Series
BI/AI/ML: Tableau, Power BI, Streamlit, Excel, Pandas, NumPy, SciPy, Plotly, scikit-learn, NLTK, Tensorflow, PyTorch
Big Data: AWS, GCP, Azure, Spark, PySpark, Databricks
DevOps: Git, GitHub, RegEx, CI/CD, Agile, Jira
Database: MySQL, PostgreSQL, NoSQL (MongoDB)

PROFESSIONAL EXPERIENCE

Oak AI, Boston, MA - *Data Scientist* ***Jan 2024 – Present***

In-house consultant at a cutting-edge ethical AI software services firm, directly reporting to the CEO

Tech Stack: Python, Agile, NLP, ML, AWS, PostgreSQL, Supabase, ChromaDB, RAG, Gen AI, ChatGPT API

- Generated sentence vector embeddings using BERT transformer model to perform cosine-based clustering of sentences from various documents to identify similarities and reduce manual review time by **40%.**
- Assigned topics to the sentence clusters using a RAG-based (Retrieval-Augmented Generation) approach, enhancing the accuracy of topic modeling by leveraging external knowledge sources
- Developed and deployed edge functions on Supabase to integrate the sentence clustering and topic assignment model into the production database
- Performed manual testing of the deployed edge functions using Postman to ensure functionality, validate API responses, and confirm seamless integration with the production database, ensuring reliable and accurate real-time text analysis
- Contributed to various NLP tasks like Named Entity Recognition (NER), Topic Modeling, and Sentiment Analysis, enhancing the system's ability to extract key information, categorize content, and interpret sentiment from large a corpus of documents

Skyworks Solutions Inc, Irvine, CA - *New Product Engineer Intern* ***Sep 2022 – Dec 2022***

Solutions specialist pioneering impactful data solutions in a leading S&P 500 semiconductor manufacturing firm.

Tech Stack: Python, Jira, Neural Networks, CNN, ML, CI/CD, GitHub, TensorFlow, Image Segmentation

- Generated a robust image segmentation model for semiconductor wafer bonding layers to process **10000+** wafer images
- Enabled efficient multiclass classification of wafers with good and bad bonding quality, distinguishable by pixel intensity.
- Implemented U-Net Convolutional Neural Network architecture using TensorFlow and Python Jupyter Notebook to efficiently segment the images, achieving an accuracy of **92%**

Robert Bosch Engineering and Business Solutions, India - *Associate Software Engineer Intern* ***Jan 2020 – Jun 2020***

Consultant in operating services team of global automotive technology services provider.

Tech Stack: Python, ML, Databricks, Spark, Tableau, MySQL

- Forecasted the efficiency of Advanced Driver Assistance System (ADAS) APIs by utilizing statistical analysis techniques and machine learning algorithms, improving predictive accuracy and system performance by **18%.**
- Assessed the performance of linear regression models using AUC and ROC curves to identify the most effective model features.
- Reported key performance indicators to senior management by building interactive Tableau dashboards.

PROJECTS

Stock Price Prediction with Sentiment Analysis – *Python, ML, NLP, Streamlit, Time Series Analysis* | Project Link

- Developed and implemented a comprehensive stock market forecasting project leveraging quantitative data analysis and sentiment analysis from social media platforms such as Twitter, Reddit, and news articles.
- Integrated advanced machine learning models, including Multivariate LSTM and Prophet, to predict future stock prices.
- Incorporated sentiment features to capture the nuanced relationship between market dynamics and public sentiments.