Shravani Nimbolkar

nimbolkarshravani.github.io | nimbolkarshravani@gmail.com | linkedin.com/in/shravaninimbolkar | https://github.com/nimbolkarshravani

SUMMARY

Hi! I am a passionate data enthusiast proficient in Python, SQL, and AI frameworks like TensorFlow and PyTorch. I am also experienced in developing end-to-end data-driven systems, building data pipelines, and fine tuning LLMs. I am committed to utilizing AI for social and environmental good by leveraging cutting-edge technologies to make a positive impact.

SKILLS

- **Languages**: Python, SQL, CPP, Frontend: HTML, CSS, JS;
- **Frameworks**: Pandas, Numpy, Sklearn, TensorFlow, Pytorch, Keras, Regression, Ensemble techniques, CatBoost, Statsmodel, Neural nets, SimpleTransformers, OpenCV, NLTK, Flask, Git, RapidMiner, Pypfopt, GCP, BERT, Pycarat, H2O, BigQueryML;
- **Dashboarding**: PowerBi, Tableau, Looker, Google analytics dashboard, MS Excel

EXPERIENCE

Data Science Intern| OmniThink.ai
May 2023 - Aug 2023
- Optimized queries in **BigQueryML** on a dataset of 14.5K rows resulting in 20% reduction in query execution time. Analyzed the data to find the anomalies and generated interactive visualization dashboards on **Looker** Studio to present actionable insights to the stakeholders.
- Implemented ARIMAX for sales forecasting and used **BERT** for sentiment analysis on product reviews to leverage the forecast. Performed statistical analysis and **hypothesis testing** for feature selection and VIF technique in **Statsmodel** to mitigate multicollinearity.
- Used hyperparameter tuning to reduce the mean error by 6% compared to baseline ARIMA PLUS model.
- Built and managed data pipelines in Dataflow to streamline analytics workflows in real-time.

Associate Analyst | Schlumberger
Aug 2021 - Jul 2022
- Maintained 98%+ uptime for 150+ applications across the organization through monitoring and troubleshooting using Dynatrace and **SAP Solman** eradicating 75% critical issues. Leveraged Power Automate to automate 50% of documentation workflows, streamlining processes and increasing operational efficiency.
- Created and deployed real-time **PowerBI** dashboards for stakeholders to track application's performance metrics and KPIs for informed decision making.
- Played a critical role during system outages by promptly reporting alerts, collaborating with SREs, product owners and application teams, and coordinating resolution efforts. Demonstrated strong collaboration skills in a high-pressure environment.

Machine Learning Intern| Centre for Development of Advanced Computing
Sep 2020 - Mar 2021
- Researched about **LLMs** and Implemented a **BERT** classifier in **SimpleTransformers** on Liar Liar (comprising 12.K rows) and PolitiFact(comprising 21K rows) dataset.
- Performed text pre-processing NLP tasks and extracted out stylistic attributes of the text like length, grammar, word frequency with a few additional features like stance, sentiment after carefully studying their correlation with target variable to refine the prediction.
- Fine-tuned the BERT model on 80% data which yielded an **accuracy of 82%**, achieved a 4% increase in accuracy compared to the baseline model.

EDUCATION

University of California, Berkeley
May 2023
MS in Analytics| GPA 3.6
Vice-president of technology |Berkeley Analytics Consulting Organization

Pune University
Jul 2021
Bachelor of Computer Engineering | CGPA: 9.0/10.0
Courses: DSA, DBMS, Web Technology, Software engineering & Project management, AI & Robotics, Data Analytics, Data mining & warehouse, Machine learning

NOTABLE PROJECTS

Lung Disease Prediction | CXRDiagnosis | 1st Prize | GitHub | Book chapter
Jun 2021
- Built a classifier called MetaEfficientNet trained on Chest X Ray image dataset having an accuracy of 97% with 3 classes; Pneumonia, COVID19 & Normal.
- Used Few shot learning approach to further train the model on limited data. Used Metric-based Siamese net consisting of two EfficientNetB0 models as sister networks to learn the image representation in feature space.
- Achieved an **accuracy of 85.5%** for 6 newly introduced classes of lung diseases; successfully deployed the model on a web app using flask and Heroku; led a team of 3 and won 1st prize among 23 project teams.

SensorAI - Human Activity Recognition | GitHub
Dec 2020
- Aggregated sensor readings from 24 subjects engaged in 6 distinct physical activities, consolidating them into a unified dataset comprising about 15,000 rows. Performed initial data preprocessing and feature engineering, subsequently refining the dataset by extracting 45 relevant attributes followed by feature scaling.
- Trained a deep neural network on 75% of the data, while also creating simulations to account for scenarios involving sensor malfunctions or absences. Built a robust model with a test **accuracy of 93.4%**.
- Utilized **Tensorflow.js** to deploy the final model in a cost-effective and highly scalable manner, eliminating the need for a backend infrastructure.

Other Projects
- PortfolioPro (Optimization) , Customer Segmentation (Clustering), Walmart Sales(Timeseries) , AI for an Eye(Computer vision), AgroML (Computer vision)

ACHIEVEMENTS

- Data Centric AI Competition 2023| Multi-class Text Classification, used ensemble of BERT, RoBERT — *Rank 23 /343 participants| Feb 2023*
- Uhack Sentiment 2.0: Decode Code words | Merkel challenge| Multi-Label Text Classification, used RoBERT, DistilBERT — *Rank 5/1041 participants | Jan 2022*
- Deloitte Machine Learning Challenge | CatBoost, XGBoost, LightGBM, Random Forest | Notebook — *Rank 44 / 1484 participants | Dec 2022*
- **Merit based scholarship:** Received recognition for outstanding academic accomplishments by UC Berkeley.
- MTA Introduction to Programming Using Python - Certified 2020 (82%) Credentials
- Coursera - ChatGPT Prompt Engineering for Developers, Understanding and Applying Text Embeddings with Vertex AI More certificates