Shravani Nimbolkar

+15105707497 | nimbolkarshravani.github.io | nimbolkarshravani@gmail.com | linkedin.com/in/shravaninimbolkar | https://github.com/nimbolkarshravani

SUMMARY

Dedicated Data Scientist with over **4 years** of experience in building and deploying Machine learning and Deep learning models with proficiency in **Python, SQL**, cloud technologies, Tableau data visualization, NLP, NLG, **docker, Kubernetes**, and variety of AI frameworks(Keras/TensorFlow/PyTorch). I specialize in extracting actionable insights from large structured and unstructured datasets.

SKILLS

- **Languages**: Python, SQL, R, C++, Frontend: HTML, CSS, JS;
- **Frameworks**: Pandas, Numpy, Sklearn, **TensorFlow**, Pytorch, Keras, Regression, Ensemble techniques, CatBoost, Statsmodel, Simple Transformers, Streamlit, OpenCV, NLTK, Flask, Git, RapidMiner, PySpark, Pypfopt, GCP, Docker, Kubernetes, H2O, BigQueryML, Langchain, Gemini Pro, ChatGPT, Prompt Engineering;
- **Dashboard and Database**: PowerBI, Tableau, SQL Workbench, Looker Studio, Google analytics dashboard, SaS, MS Excel
- **Certification**: MTS Introduction to Python, Google Analytics, Coursera: Deep Learning Specialization, Artificial Intelligence: **Ethics** & Societal Challenges etc

EXPERIENCE

Data Scientist | Biomotivate | Pittsburgh, USA   *Oct 2023 - Present*
- Designing and fine-tuning **Deep Learning models** to predict early treatment dropouts using physiological data from patients at a rehabilitation centre.
- Utilizing **PowerBI** for data analysis and Python to employ time series analysis techniques to fine-tune 'Empatica' E4 Electro-Dermal Activity (EDA) data at different frequencies. Led data dives to analyze Empatica E4 Electro-Dermal Activity (EDA) data, conducting in-depth hypothesis testing to gain a nuanced understanding of physiological responses for improving model's performance.

Data Science Intern| OmniThink.ai | San Francisco, USA   *May 2023 - Aug 2023*
- Optimized **SQL** queries in BigQueryML on a dataset of 14.5K rows resulting in 20% reduction in query execution time. Trained and deployed ARIMA PLUS and ARIMAX timeseries model on **GCP Cloud environment**. Developed interactive visualization dashboards on **Looker Studio** to convey actionable insights to stakeholders.
- Used BERT/GPT **API** 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 | Pune, India   *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 | Mumbai, India   *Sep 2020 - Mar 2021*
- **Researched** about **LLMs** and Implemented a **BERT** classifier in **Simple Transformers** 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 along with like stance detection, emotion and **toxicity** analysis and studied their correlation with the 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   *Aug 2023*
MS in Analytics| GPA 3.6 |**Vice-President of Technology** @ Analytics Consulting Organization at Berkeley

Pune University   *Jul 2021*
Bachelor of Computer Engineering | CGPA: 9.0/10.0
Courses: DSA, Database Management, Software engineering & Project management, AI & Robotics, Data Analytics, Data Mining & Warehouse, Machine Learning

NOTABLE PROJECTS

Chat With PDF| GitHub   *Jan 2024*
- Developed a **Chatbot** for interactive querying of information from multiple PDF documents using Google Gemini Pro and LangChain. Incorporated **NLG** techniques to enhance the conversational aspect of the Chatbot, providing users with natural humanlike responses.
- Implemented efficient text **summarization** and indexing with Google Generative AI **Embeddings**, ensuring accurate information retrieval from large datasets.
- Designed and implemented a streamlined web application using Streamlit, enhancing the user experience for asking questions and receiving detailed responses.
- Leveraged Python, Streamlit, PyPDF2, LangChain, Google Generative AI, and FAISS to create an innovative solution for conversational PDF interaction.

Analysing Customer Reviews |Uhack Sentiment 2.0: Decode Code words | Rank 5   *Jan 2022*
- Analyzed customer review dataset consisted of 6000+ rows and 12 different labels about product usability, functionality, installation, design and aesthetics etc.
- Performed text pre-processing **NLP** techniques and fine-tuned an ensemble of 8 RoBERTa models for multi-label classification
- Used multi-label stratified cross-validation for hyperparameters tuning and minimizing the logloss.
- Ranked 5th on the leaderboard (logloss=2.781), demonstrating the effectiveness of the **Transformer** ensemble approach.

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 as sister networks and head model **CNN** to learn the image **embeddings** in feature space.
- Achieved an accuracy of 85.5% for 6 newly introduced classes of lung diseases; successfully **docker**-containerized the model and deployed web app using flask and **Salesforce Heroku**; led a team of 3 and won 1st prize among 23 project teams.

DELOITTE Machine Learning Challenge |Risky Loan Defaulter Prediction | Top 3%| Notebook   *Jan 2021*
- Developed an ensemble of CatBoost,Light GBM and XGBoost trained on a dataset of 67k rows, 35 features to predict Loan defaulters to achieve log loss of 0.36.
- Performed feature selection and used SMOTE, KNN and Down sampling to handle class imbalances to further reduce the loss to 0.34
- Finally performed error analysis to understand the model and quantized risk using Conformal Prediction

OTHER PROJECTS

- AgroML (Crop Weed Detection using AI)
- Sensor AI (Human activity recognition using Smart phone sensors)
- Customer Segmentation (Consumer behaviour Analysis)

HACKATHONS

- **GOOGLE** |BigQuery ML Data Centric AI Competition | Feb 2023| Rank 23
- **ZS**| ACE – A – THON | Top 10
- INNOVATE FOR SMART PCMC | 2nd Prize