PRATIKSHA PRADEEP SHARMA

(352)-870-6364 | pratikshas2343@gmail.com| Github | LinkedIn

EDUCATION

Master of Science, Agriculture, and Biological Engineering | University of Florida **Aug 2021 - Dec 2023**

Master of Science, Electrical and Computer Engineering | University of Florida **Aug 2019 - Dec 2021**

Bachelor of Science, Electronics and Telecommunications Engineering | University of Mumbai **Jul 2015 - May 2019**

SKILLS

- **Languages:** Python, SQL, R, Matlab
- **Tools and Software:** Jupyter Notebook, Anaconda, Tableau, Apache Spark, MS Excel, AWS, Git
- **Libraries:** NumPy, Pandas, Tensorflow, Keras, Matplotlib, Seaborn, BeautifulSoup, PyTorch, Plotly, Scikit-Learn, Scipy, NLTK, ggplot2, dplyr, StatsModels
- Machine Learning, Deep Learning, Natural Language Processing
- Data Mining, Pattern Recognition, Statistical Analysis

EXPERIENCE

Graduate Research Assistant | University of Florida **Mar 2021 – May 2023**

- Estimated the probability of pregnancy of cows using ML models (**Logistic Regression, XGBoost, Random Forest, and MLP)** for an imbalanced dataset containing 29% positive outcomes of insemination
- Conducted extensive EDA on the activity, and health information of 26,000 cow records to identify trends, and factors influencing pregnancy
- Outlier handling through winsorization, feature selection using Gini Importance, and feature engineering enhanced the predictive power of the models.
- Incorporated Hyperparameter tuning and 10-fold cross-validation on the models
- The best-performing model was XGboost which attained an AUC of 0.75 and an F1-score of 0.47

PROJECTS

Regeneration of Images and text using generative learning and deep neural networks

- Developed and implemented GANs and VAEs to generate digits using the MNIST dataset
- Compared the results of a CNN classification model on the generated digits with manual labeling, achieving a 76% accuracy by the GAN model
- Trained an LSTM model on 50,000 reviews from the IMDB movie review data and generated coherent and related sentences by providing a seed

Bollywood Movie Recommender System

- Implemented a content-based recommender system on the TMDB 5000 movies dataset using cosine similarity
- Leveraged natural language processing techniques (CountVectorizer) to represent significant movie contents
- Created an interactive web application using Streamlit with Heroku deployment

Customer Analytics for Business Growth

- Employed Hierarchical and K-means clustering techniques and PCA to reduce dimensionality and segment customers
- Conducted descriptive analysis by brand and segment to identify customer behavior patterns
- Used linear and logistic regression models for elasticity modeling, including purchase probability, brand choice, and purchase quantity
- Prediction of future customer behavior by an Artificial neural network achieved an accuracy of over 90%