SANDEEP UNDURTHI

+1 (510) 980 1853   - sandy.undurthi@gmail.com

EDUCATION

Utah State University, Logan, Utah, USA
 Master's of Computer Science

TECHNICAL SKILLS

- **Languages and Frameworks:** Python, SQL, Java, C
- **Data Tools:** MySQL, Excel
- **ML and NLP Libraries:** TensorFlow, PyTorch, Scikit-learn, Numpy, Pandas, Matplotlib, BERT, K-Means
- **Development Environment:** Jupyter Notebook, VS Code

PROJECTS

Fraud Detection Using Machine Learning & Deep Learning

Developed a comprehensive fraud detection system to identify fraudulent credit card transactions with high accuracy.

Key Achievements:

- Preprocessed and analyzed a highly imbalanced dataset using Pandas, NumPy, and Seaborn to detect fraud patterns
- Trained and compared ML models (Logistic Regression, Random Forest, XGBoost) achieving an AUC-ROC score of 0.99
- Implemented an Autoencoder for anomaly detection, reducing false positives using reconstruction errors
- Evaluated models using Confusion Matrix, Precision-Recall, and ROC Curve Analysis
- Utilized TensorFlow/Keras to train deep learning models for detecting fraudulent transactions
- Visualized model performance with Matplotlib and Seaborn to interpret fraud patterns effectively

House Price Prediction Using Regression Models

Developed a machine learning model to predict house prices based on real estate features including location, size, bedroom/bathroom count, and construction year.

Key Achievements:

- **Data Preprocessing & Feature Engineering:**
  - Cleaned and handled missing data using median imputation for numerical features and mode imputation for categorical features
  - Applied One-Hot Encoding to convert categorical variables into numerical form
  - Removed outliers to improve model robustness
- **Model Training & Evaluation:**
  - Implemented multiple regression models and evaluated using Root Mean Squared Error (RMSE):
    - Linear Regression RMSE: 44,830
    - Ridge Regression RMSE: 24,164
    - XGBoost RMSE: 22,019 (Best Model)
  - Optimized performance by fine-tuning hyperparameters
- **Deployment & Results:**
  - Selected XGBoost as the final model for making accurate price predictions
  - Achieved a score of 0.14 on the Kaggle House Prices Competition leaderboard

Twitter Sentiment Analysis

Analyzed tweet sentiment at the entity level, determining whether messages express Positive, Negative, or Neutral sentiment toward specific entities.

Key Achievements:

- **Data Preprocessing & Feature Engineering:**
  - Cleaned and preprocessed tweet text
  - Utilized TF-IDF Vectorization to convert text into numerical format
  - Handled missing values and standardized sentiment labels
- **Model Development & Comparison:**
  - Implemented and compared multiple models with the following accuracy results:
    - Naive Bayes: 71.37%
    - Logistic Regression: 81.88%
    - Decision Tree: 88.69%
    - Random Forest: 93.89%
    - Artificial Neural Network (ANN): 96.89% (Best Model)
  - Demonstrated the effectiveness of deep learning for text classification and NLP tasks