Sanny Un Sowadh Wamik

sannyunsowadh01@gmail.com Melbourne, VIC sanny-un-sowadh-wamik Sanny-Un-Sowadh-Wamik

EDUCATION

Bachelor of Data Science 06/2027
Royal Melbourne Institute of Technology Melbourne, Australia
Relevant Coursework: Practical Data Science, Advanced Programming for Data Science, Data Visualisation with R, Foundations of Artificial Intelligence, Practical Statistics, Object-Oriented Programming, Designing

PROJECTS

Music Bot
Team Lead
MusicBot: Discord music streaming bot built with **Wavelink API** and **Lavalink** integration. Features high-quality audio playback, playlist management, and intuitive commands for seamless server entertainment.

Depression Prediction Project
Competed the 2024 Kaggle Playground Series – Season 4, Episode 11 "Exploring Mental Health Data" competition, achieving a **0.94125** ROC-AUC score on the test set. Leveraged Python (Pandas, NumPy), Scikit-Learn pipelines and transformers (SimpleImputer, OneHotEncoder, StandardScaler), and LightGBM classifier within a stratified cross-validation framework. Employed EDA and statistical analysis (**chi-square tests**, **t-tests**), extensive feature engineering, hyperparameter optimization via RandomizedSearchCV, and ensemble techniques to identify key predictors of depression. Core skills and keywords to feature: Machine Learning, Predictive Modeling, Ensemble Learning, Feature Engineering, Hyperparameter Tuning, Data Preprocessing, Model Evaluation, Cross-Validation, Statistical Analysis, Data Visualization, Python, Scikit-Learn, LightGBM, Kaggle Competition Experience.

Data Preprocessing & Integration Pipeline | RMIT University
Developed a comprehensive data pipeline for e-commerce clothing reviews, implementing **XML parsing**, data cleansing, and multi-source integration using Python. Engineered solutions for schema conflicts, data inconsistencies, and quality issues using pandas, ElementTree, and statistical analysis. Demonstrated strong data wrangling abilities with automated error detection and documentation, producing a unified dataset ready for analytics applications.

Titanic - Machine Learning from Disaster
Developed a end-to-end machine learning pipeline to predict passenger survival on the Titanic (binary classification) using Python. Performed data exploration, cleaning and feature engineering (handling missing values, encoding categorical variables, creating interaction terms), then trained and tuned models (e.g. Random Forest, Logistic Regression) to optimize performance. Achieved a Kaggle score of **0.76794** (76.794% accuracy) through cross-validation and threshold calibration.

House Prices Prediction
Developed an end-to-end regression pipeline for the Kaggle "House Prices: Advanced Regression Techniques" competition, leveraging the Ames dataset with **79** features. Performed extensive data preprocessing and feature engineering (missing-value imputation, log-transformation of SalePrice, encoding of categorical variables, outlier treatment), then trained and fine-tuned multiple models—including Lasso, Ridge, ElasticNet, Random Forest, Gradient Boosting, XGBoost, and LightGBM—using 5-fold cross-validation (RMSE). Combined predictions via simple averaging to achieve a public leaderboard RMSE of **0.12099**.

ACTIVITIES AND LEADERSHIP

RMIT CSIT Society 08/2024 – present
CSIT Insider Melbourne, Australia

Notre Dame International Understanding and Relation Club 2021 – 2023
Organizer Dhaka, Bangladesh

SKILLS

Programming: Python (Pandas, NumPy, Matplotlib), R (ggplot2, dplyr), SQL
Tools: Jupyter Notebook, RStudio, Git, PyCharm, VsCode.
Data Science: Statistical analysis, data visualization, exploratory data analysis, machine learning fundamentals
Cyber Security: Data protection principles, basic security protocols