Data Science/Machine Learning Portfolio

Hang Pham
hang.pth191@gmail.com
github.com/HangPham91
linkedin.com/in/hangpham91

©2025☆HangPham☆

Overview

Purpose: Showcase applied machine learning and data analysis projects across diverse domains.

Domains Covered:

- Finance (stock forecasting)
- Automotive (used car prices)
- Disease diagnosis (Parkinson's)
- Women's health (menstrual cycles)
- Medical imaging (breast cancer)

Techniques Used:

- Data cleaning, visualization, feature engineering
- Machine Learning Regression, classification, deep learning, time series forecasting
- Model evaluation and interpretation

Featured Projects

- Tesla Stock Price Prediction using LSTMs model
- Parkinson's disease classification with Machine Learning Classifiers
- Breast Cancer Classification with Convolutional Neural Networks
- Menstrual cycle analysis and prediction with Machine Learning Regressors
- Ford car price analysis and prediction with Machine Learning Regressors

©2025☆RangPham☆


Tesla Stock Price Forecasting

Goal:

- Forecast Tesla's next-day closing price using historical data and deep learning model Long Short Term Memory (LSTM)

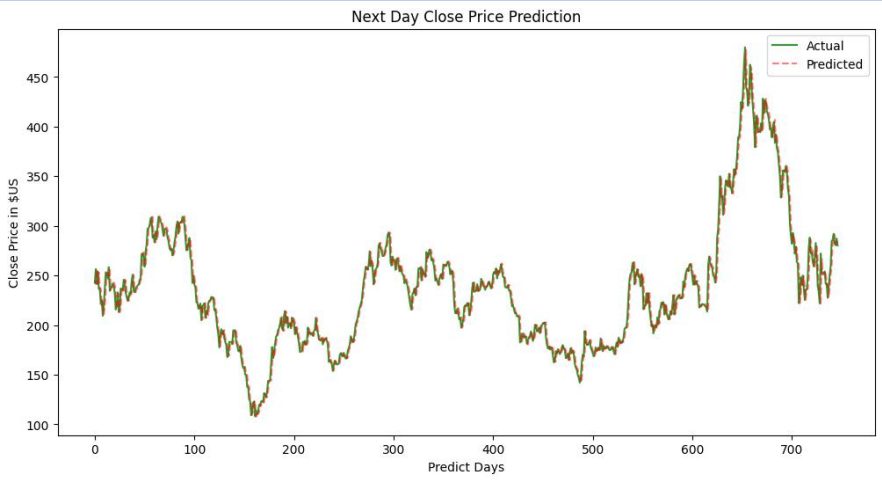
Tools:

- Python, pandas, numpy, yfinance, matplotlib, seaborn, plotly
- Scikit-learn: MinMaxScaler, metrics
- TensorFlow/Keras: LSTM neural networks
- Linear Regression (baseline)

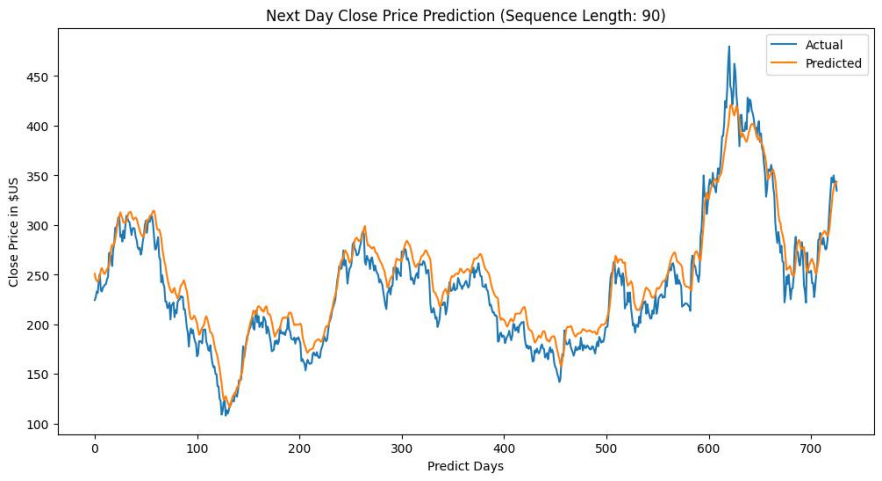
Key Results:

- LSTM models (30, 60, 90, 120 days) outperformed Linear Regression
- Results summarized and visualized for comparison
- The best LSTM model achieved the lowest **MAPE of 5.02%** and the highest **R^2 up to 0.94**

Linear Regression model



LSTM model (90days)



Parkinson's disease classification

Goal:

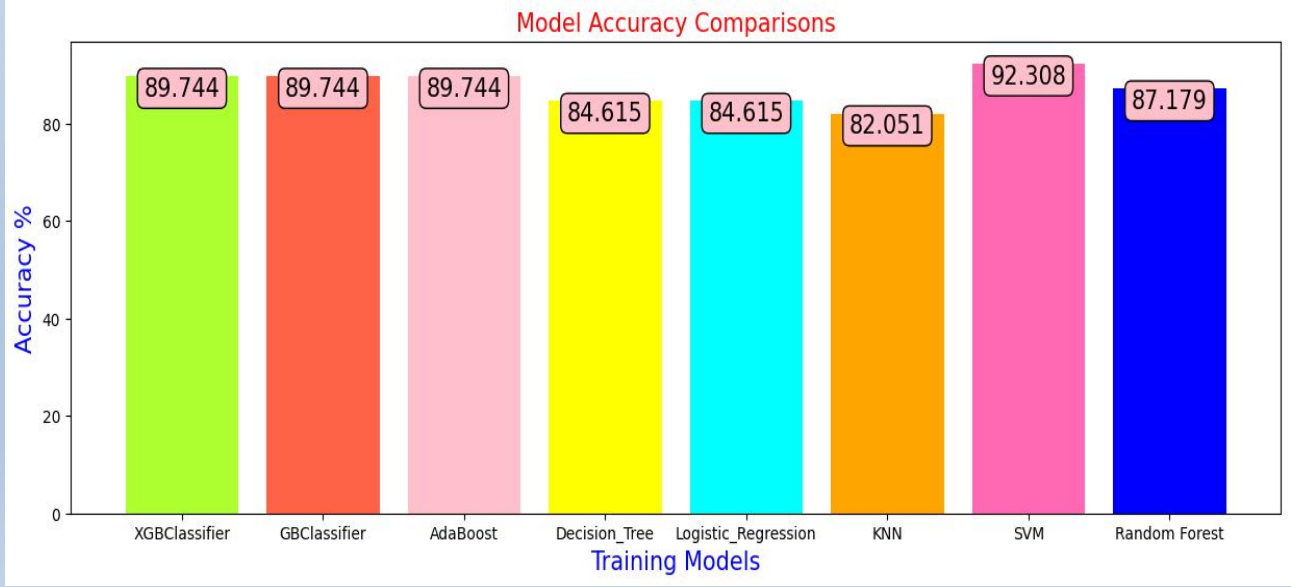
- Classify patients as healthy or with Parkinson's disease using voice features

Tools:

- Python, pandas, numpy, seaborn, matplotlib
- Scikit-learn: Logistic Regression, SVM, KNN, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, XGBoost

Key Results:

- Feature distributions and confusion matrices provided insights into the dataset
- Benchmark multiple classifiers, average **accuracy** rates ranged from **82% to 92%**
- **SVM** achieved the **highest accuracy** among all classifiers, e.g., 92%



Breast Cancer Classification (CNN on Mammograms)

Goal:

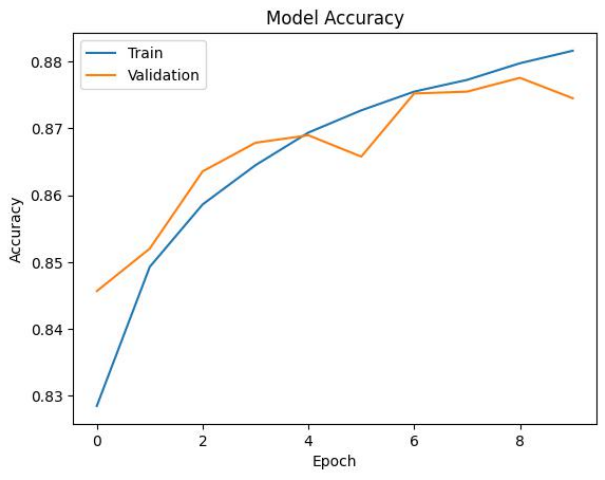
- Detect and classify breast cancer from mammogram images

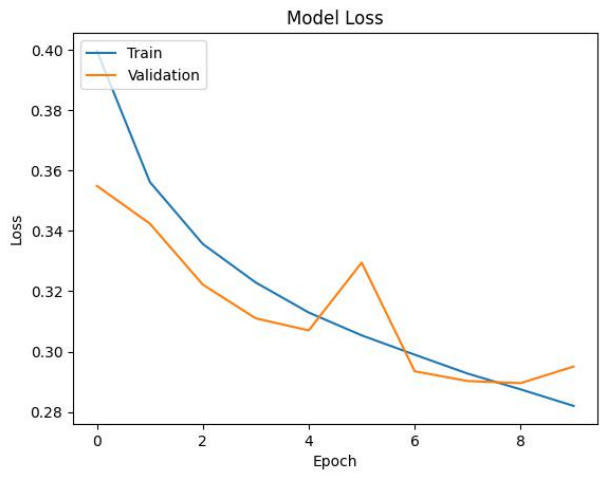
Tools:

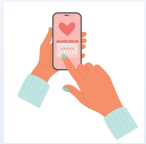
- Python, pandas, numpy, matplotlib, seaborn
- TensorFlow/Keras: CNN (Conv2D, MaxPooling2D, Dense, Dropout)
- ImageDataGenerator

Key Results:

- Built a 4-layer CNN for IDC vs. non-IDC classification
- Visualized model performance (accuracy/loss plots)
- Achieved high training and validation accuracy of 87%







Menstrual Cycle Analysis & Prediction

Goal:

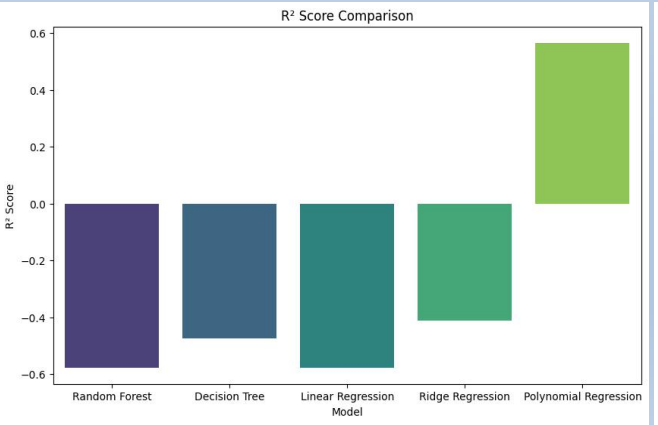
- Classify menstrual cycles as regular/irregular
- Analyze the impact of stress and age on the menstrual cycle
- Predict the next cycle length using Machine Learning

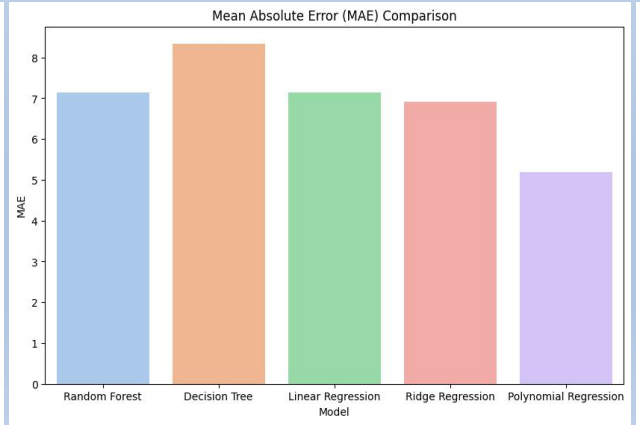
Tools:

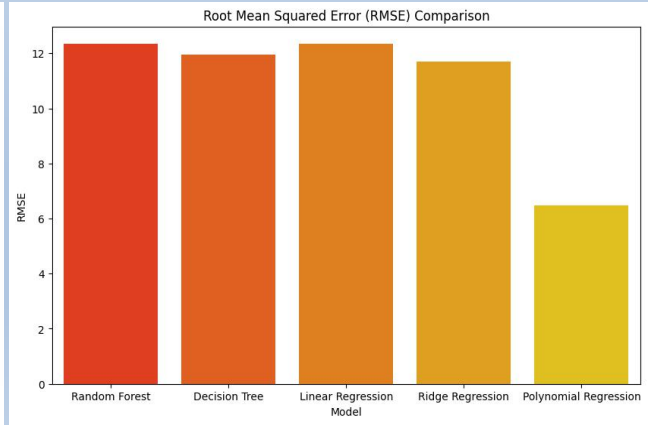
- Python: pandas, numpy, seaborn, matplotlib, plotly
- Scikit-learn: Linear Regression, Random Forest, Decision Tree, Ridge, Polynomial Regression

Key Results:

- High stress and age linked to longer/irregular cycles
- Polynomial Regression achieved the best prediction, e.g., lowest MAE, RMSE and highest R^2
- Provided actionable prediction for next cycle start date







Ford used Car Price Analysis & Prediction

Goal:

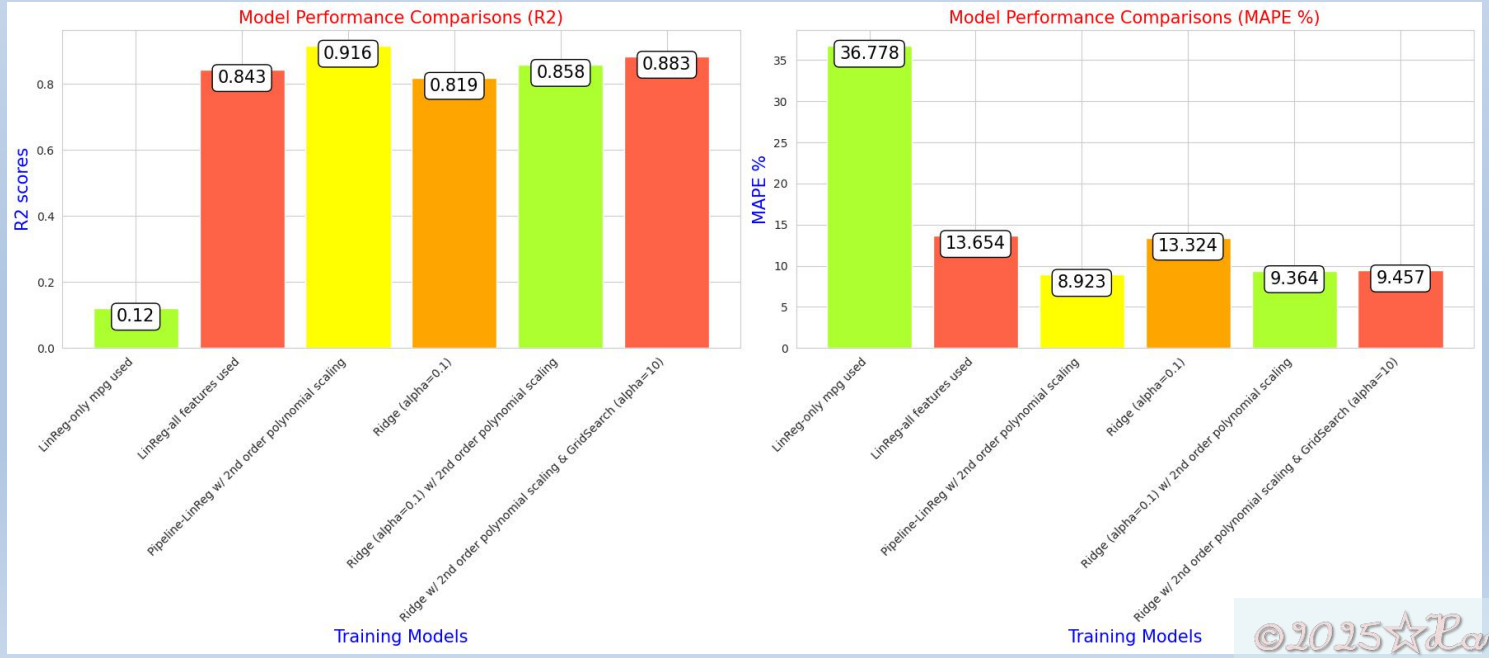
- Analyze and predict used Ford car prices based on features

Tools:

- Python, pandas, seaborn, matplotlib
- Scikit-learn: Linear Regression, Ridge, Polynomial Features, Pipeline, GridSearchCV

Key Results:

- Visualized feature importance and model performance
- The pipeline model of StandardScaler with Linear Regression and 2nd-order Polynomial Features yielded the best R^2 of 92% and the lowest MAPE of 8.9%



Summary

- Diverse Applications: Finance, Health, Automotive, Medical imaging, Disease diagnosis, Women's Health
- Comprehensive Workflow: how I complete a project

Problem definition
(choose and understand the field/business/question of interest & define the goal)

↓

Data Collection (e.g, Kaggle datasets, Yahoo Finance...)

↓

Data Cleaning and Preprocessing
(Handle missing values, outliers, duplicates, encoding, scaling)

↓

Exploratory Data Analysis (EDA)
(e.g. Data visualization, uncovering patterns, trends, correlations)

↓

Feature Engineering

↓

Model Selection
(Choose appropriate algorithms based on the problem type)

↓

Model evaluation
(Assess model performance using metrics)

- Modern ML Techniques: Deep learning, ensemble methods, regression, classification
- Actionable Insights: Each project delivers interpretable results and practical recommendations

Contact

I'm currently seeking full-time opportunities in Data Science, Data Analysis, Machine Learning Engineering, and Quantitative Research. Let’s connect! 🤝

hang.pth191@gmail.com

linkedin.com/in/hangpham91/

github.com/HangPham91

©2025☆HangPham☆