**BigMart Sales Prediction Project Documentation**

**Project Overview**

The BigMart Sales Prediction project aims to forecast sales for various products at different store locations. Using advanced machine learning models, we predict sales based on historical data, enabling better inventory management and business decision-making.

**Dataset Description**

**Files Used:**

* train.csv: Training dataset containing features and sales data.
* test.csv: Testing dataset without sales figures.

**Features:**

1. **Item\_Identifier:** Unique product code.
2. **Item\_Weight:** Product weight.
3. **Item\_Fat\_Content:** Fat content (low fat or regular).
4. **Item\_Visibility:** Product visibility in the store.
5. **Item\_Type:** Product category.
6. **Outlet\_Identifier:** Store identifier.
7. **Outlet\_Establishment\_Year:** Store establishment year.
8. **Outlet\_Size:** Store size.
9. **Outlet\_Location\_Type:** Store location type.
10. **Outlet\_Type:** Store type.
11. **Item\_Outlet\_Sales:** Target variable in the training dataset.

**Data Preprocessing**

**1. Missing Value Treatment**

* Filled missing Item\_Weight with the mean.
* Filled missing Outlet\_Size with 'Unknown'.

**2. Feature Engineering**

* Created Item\_Age by subtracting Outlet\_Establishment\_Year from the current year.
* Created log and squared transformations for Item\_Visibility.
* Created interaction features such as Item\_Weight\_Visibility.

**3. Outlier Removal**

* Removed outliers using the Interquartile Range (IQR) method.

**4. Encoding and Scaling**

* Applied OneHotEncoding for categorical features.
* Used QuantileTransformer for numerical scaling.

**Model Development**

**1. Baseline Models**

* **XGBoost:** Tuned using GridSearchCV with RMSE scoring.
* **Random Forest Regressor:** Ensemble learning model.
* **Ridge Regression:** Linear regression with L2 regularization.

**2. Advanced Models**

* **Stacking Regressor:** Combining XGBoost, Random Forest, and Ridge models.
* **Voting Regressor:** Ensemble averaging from multiple models.

**3. Neural Network Model**

* Built a deep learning model using TensorFlow with different configurations:
  + Variable hidden layers, neurons, activations, optimizers, and dropout rates.
  + Performed grid search to find optimal model parameters.

**Model Evaluation**

* Evaluation Metric: Root Mean Squared Error (RMSE)
* Best Validation RMSE Achieved: Recorded for each model.

**Predictions and Submission**

* Selected the best model based on RMSE.
* Predicted sales on the test dataset.
* Created multiple submission files: submission\_P.csv, submissionm2.csv, submissionm3.csv, and submission\_ann.csv.

**Future Enhancements**

* Experiment with additional models like LightGBM and CatBoost.
* Implement more complex feature selection methods.
* Use deep learning architectures like LSTMs or Transformers for sequential sales forecasting.

**Conclusion**

This project successfully built a predictive model for BigMart sales forecasting using machine learning and deep learning models. Through careful feature engineering, model selection, and hyperparameter tuning, we achieved significant predictive accuracy.