

Change Logs for Machine Learning Algorithms

Each table provides details on:

- What changes were made
- Why those changes were made
- Time taken (in hours)
- Difficulty rating (out of 10)
- Whether the same notebook was used or a new one was created

Random Forest Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Improved feature selection	To reduce overfitting	3.5	7	Modified
Tuned hyperparameters	To increase model accuracy	4.5	8	Modified
Added cross-validation	To ensure better generalization	3.0	6	New Notebook
Optimized tree depth	To improve computational efficiency	2.5	7	Modified

Total Hours: 13.5

SVM Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Implemented GridSearchCV	To find optimal hyperparameters	3.5	7	Modified

Normalized dataset	To improve model performance	2.0	5	Modified
Switched to RBF kernel	To handle non-linearly separable data	3.0	6	New Notebook
Adjusted regularization parameter	To prevent overfitting	2.5	7	Modified

Total Hours: 11.0

Naïve Bayes Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Preprocessed text data	To make the model compatible with text	3.0	6	Modified
Added Laplace smoothing	To prevent zero probability issues	2.0	5	Modified
Used TF-IDF vectorization	To improve text classification	3.5	7	New Notebook
Implemented multinomial Naïve Bayes	To enhance classification for text data	2.5	6	Modified

Total Hours: 11.0

Decision Tree Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Handled missing values	To avoid errors in training	2.5	5	Modified
Pruned decision tree	To prevent overfitting	3.0	7	Modified
Switched to entropy	To improve accuracy	3.5	6	New Notebook

Added max-depth tuning	To optimize tree complexity	2.0	6	Modified
------------------------	-----------------------------	-----	---	----------

Total Hours: 11.0

K-Means Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Found optimal K using Elbow Method	To determine the best number of clusters	2.5	6	Modified
Standardized features	To improve clustering accuracy	1.5	5	Modified
Implemented silhouette analysis	To validate clustering results	3.0	7	New Notebook
Added PCA for dimensionality reduction	To improve clustering efficiency	2.0	6	Modified

Total Hours: 9.0

KNN Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Tried different K values	To find the optimal neighbors	2.5	6	Modified
Normalized dataset	To improve distance calculations	2.0	5	Modified
Implemented weighted KNN	To improve classification results	3.5	7	New Notebook
Optimized distance metric (Euclidean vs Manhattan)	To enhance accuracy	2.0	6	Modified

Total Hours: 10.0

CNN Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Built CNN baseline model	To establish a performance benchmark	3.0	6	New Notebook
Visualized sample training images	To understand dataset distribution	0.5	3	Modified
Normalized pixel values	To ensure consistent scale and faster convergence	0.25	2	Modified
Converted labels to one-hot encoding	To prepare data for classification	0.25	2	Modified
Grayscaled 20% of training images	To analyze effect of reduced color information on performance	1.0	5	Modified
Tuned CNN with dropout and filters	To improve generalization and prevent overfitting	4.0	8	Modified
Evaluated baseline and tuned models	To compare performance and identify improvement areas	0.5	4	Modified
Visualized training accuracy	To understand learning patterns	0.5	3	Modified
Saved tuned model and predictor script	To enable deployment and sample predictions	0.75	5	Modified
Developed Flask web app	To create an interface for predictions	2.5	7	New Notebook

Total Hours: 13.25

ANN Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status

Loaded and verified data	To ensure correctness and consistency of input	0.5	3	New Notebook
Encoded categorical variables	To prepare categorical data for model input	0.5	4	Modified
Split data into training, validation, and test sets	To create separate sets for model training and evaluation	0.25	3	Modified
Standardized features using StandardScaler	To ensure consistent scale for optimal training	0.25	3	Modified
Defined model creation function	To create reusable and consistent ANN models	0.5	5	Modified
Trained models with different regularization techniques	To evaluate the impact of regularization methods	1.0	7	Modified
Displayed test accuracy and ROC-AUC for different regularization techniques	To compare model performance	0.5	5	Modified
Trained multi-branch model	To experiment with a more complex architecture	0.75	6	Modified
Evaluated multi-branch model on test set	To assess generalization capability	0.25	5	Modified
Trained model with Adam_Decay optimizer	To test optimizer impact on performance	0.5	5	Modified
Displayed optimizer comparison results	To visualize differences in optimizer performance	0.25	4	Modified
Displayed optimizer comparison visualizations	To enhance understanding through visual analysis	0.25	4	Modified

Displayed test accuracy and ROC-AUC for different optimizers	To identify best optimizer	0.25	5	Modified
Saved training history plot for Baseline model	To document baseline model training	0.25	3	Modified
Saved model, scaler, and label encoders for deployment	To prepare models for production	0.5	4	Modified

Total Hours: 6.25

LSTM Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Business Understanding defined	To establish project objectives and relevance	0.25	2	New Notebook
Loaded and verified data	To ensure correctness and integrity of the data	0.25	3	Modified
Checked for missing values	To handle missing data appropriately	0.25	3	Modified
Completed feature engineering	To enhance model performance	0.5	5	Modified
Encoded categorical variables	To prepare categorical data for model input	0.25	4	Modified
Visualized data and target variable	To understand data distribution and trends	0.25	3	Modified
Selected features and target	To identify relevant data for model training	0.25	3	Modified
Normalized numerical features	To ensure consistent scale for optimal training	0.25	3	Modified
Created sequences for time-series data	To prepare input format required for LSTM	0.5	6	Modified

Split data into training and testing sets	To evaluate model performance effectively	0.25	3	Modified
Trained baseline LSTM with 64 units	To establish a baseline for performance comparison	0.75	6	Modified
Evaluated baseline LSTM performance	To measure initial model accuracy	0.25	4	Modified
Trained tuned LSTM with 128/64 units, dropout	To optimize model performance	1.0	7	Modified
Evaluated tuned LSTM performance	To compare performance improvements	0.25	4	Modified
Visualized training history	To analyze training behavior and convergence	0.25	3	Modified
Analyzed model results	To compare and interpret baseline and tuned model results	0.5	5	Modified
Saved tuned LSTM model	To prepare for deployment	0.25	3	Modified
Saved scaler for deployment	To ensure consistent data transformation in production	0.25	3	Modified
Tested sample prediction	To verify model performance with real data	0.25	4	Modified
Included submission files	To ensure all required files for submission	0.25	2	Completed

Total Hours: 7.5

Transformer Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Defined business understanding	To define the objective of predicting stock prices	0.25	3	New Notebook

Loaded and verified data	To ensure data quality and readiness for processing	0.5	3	New Notebook
Checked dataset summary and missing values	To identify potential issues or gaps in the data	0.5	3	Modified
Completed feature engineering and encoding	To prepare features for Transformer model input	0.75	5	Modified
Generated time-series plot of closing prices	To visualize historical trends	0.25	3	Modified
Generated feature correlation heatmap	To identify relationships between features	0.25	3	Modified
Generated distribution plot of closing prices	To understand price distribution	0.25	3	Modified
Generated scatter plot of volume vs. closing price	To analyze volume-price relationship	0.25	3	Modified
Created 990 sequences of length 10 with selected features	To prepare data for Transformer input	0.5	5	Modified
Split data into training and testing samples	To create separate sets for training and evaluation	0.25	3	Modified
Built Transformer model with 2 layers, 4 heads	To build a model capable of learning time-series patterns	0.75	7	Modified
Evaluated Transformer model performance	To assess accuracy using metrics like MSE, MAE, RMSE, R2	0.5	6	Modified
Generated training and validation loss plot	To visualize model performance over epochs	0.25	3	Modified
Generated MAE plot	To analyze mean absolute error trends	0.25	3	Modified
Generated predictions for visualization	To compare actual vs predicted values	0.25	4	Modified

Generated actual vs predicted prices plot	To evaluate prediction performance visually	0.25	3	Modified
Generated prediction error distribution plot	To assess error distribution of predictions	0.25	3	Modified
Saved Transformer model and scaler	To prepare models for deployment	0.25	4	Modified
Generated sample prediction	To validate the Transformer model with real data	0.25	3	Modified

Total Hours: 6.25

ANN Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Loaded and verified data	To ensure correctness and consistency of input	0.5	3	New Notebook
Encoded categorical variables	To prepare categorical data for model input	0.5	4	Modified
Split data into training, validation, and test sets	To create separate sets for model training and evaluation	0.25	3	Modified
Standardized features using StandardScaler	To ensure consistent scale for optimal training	0.25	3	Modified
Defined model creation function	To create reusable and consistent ANN models	0.5	5	Modified
Trained models with different regularization techniques	To evaluate the impact of regularization methods	1.0	7	Modified
Displayed test accuracy and ROC-AUC for different regularization techniques	To compare model performance	0.5	5	Modified

Trained multi-branch model	To experiment with a more complex architecture	0.75	6	Modified
Evaluated multi-branch model on test set	To assess generalization capability	0.25	5	Modified
Trained model with Adam_Decay optimizer	To test optimizer impact on performance	0.5	5	Modified
Displayed optimizer comparison results	To visualize differences in optimizer performance	0.25	4	Modified
Displayed optimizer comparison visualizations	To enhance understanding through visual analysis	0.25	4	Modified
Displayed test accuracy and ROC-AUC for different optimizers	To identify best optimizer	0.25	5	Modified
Saved training history plot for Baseline model	To document baseline model training	0.25	3	Modified
Saved model, scaler, and label encoders for deployment	To prepare models for production	0.5	4	Modified

Total Hours: 6.25

RNN Log

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Loaded and verified data	To ensure correctness and consistency of input	0.25	3	New Notebook
Removed missing values from the dataset	To clean data and prepare it for modeling	0.5	4	Modified
Visualized distribution of IMDb ratings	To understand target variable distribution	0.25	3	Modified

Visualized overview length and correlation with IMDb ratings	To identify potential features	0.5	4	Modified
Performed text preprocessing and tokenization	To prepare text for RNN modeling	0.5	5	Modified
Normalized IMDb ratings and split data	To prepare for training and testing	0.25	3	Modified
Trained baseline RNN with SimpleRNN	To establish baseline performance	1.0	6	Modified
Evaluated baseline RNN performance	To measure model accuracy and errors	0.5	5	Modified
Trained tuned RNN with LSTM and dropout	To improve performance and prevent overfitting	1.0	7	Modified
Evaluated tuned RNN performance	To compare improvement with baseline model	0.5	5	Modified
Visualized training MAE and loss curves	To analyze model performance visually	0.25	4	Modified
Analyzed model improvements	To assess effectiveness of tuning	0.25	4	Modified
Saved tuned model and tokenizer for deployment	To prepare for real-world predictions	0.25	3	Modified
Displayed sample prediction	To demonstrate deployment and prediction capability	0.25	4	Modified
Included submission files	To package all necessary files for submission	0.25	3	Modified

Total Hours: 6.0

Linear Regression Model logs

Change Made	Reason for Change	Time Taken (hrs)	Difficulty (10)	Notebook Status
Loaded and verified data	To ensure correctness and consistency of input	0.25	3	New Notebook
Checked for missing and duplicate values	To clean and prepare data	0.25	3	Modified
Visualized data distribution	To understand feature-target relationships	0.3	4	Modified
Encoded categorical variables	To prepare non-numeric data for modeling	0.5	5	Modified
Split data into training and test sets	To train and validate the model	0.25	3	Modified
Scaled and normalized features	To ensure uniform data range	0.25	4	Modified
Trained Linear Regression model	To fit the model to the training data	0.5	5	Modified
Evaluated model on test data	To assess model performance	0.3	4	Modified
Visualized residuals and errors	To analyze prediction errors visually	0.25	4	Modified
Tuned hyperparameters (if applicable)	To optimize model performance	0.5	6	Modified
Analyzed model performance (R^2 , MAE, MSE)	To compare predictions with actual values	0.25	4	Modified
Saved final model	To preserve the trained model for future use	0.1	2	Modified

Generated final report and summary	To document findings and results	0.25	3	Modified
------------------------------------	----------------------------------	------	---	----------

Total Hours: 3.9

Final Summary

Algorithm	Total Hours Spent
Random Forest	13.5 hrs
SVM	11.0 hrs
Naïve Bayes	11.0 hrs
Decision Tree	11.0 hrs
K-Means	9.0 hrs
KNN	10.0 hrs
ANN	6.25
CNN	13.25
RNN	6.0
LSTM	7.5
Transformer	6.25
Linear regression	3.9
Total	108.65 hrs