

Model Development Phase Template

Date	19 June 2025
Team ID	SWTID1750052396
Project Title	Analysis of medium app reviews from google play store
Maximum Marks	10 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include a summary and training and validation performance metrics for multiple models, presented through respective screenshots.

Initial Model Training Code (5 marks):

KNN Model

```
knn_model = KNeighborsClassifier(n_neighbors=5)
knn_model.fit(X_train, y_train)
y_pred_knn = knn_model.predict(X_test)
print("KNN Classification Report:\n", classification_report(y_test, y_pred_knn))
```

Naïve Bayes Model

```
nb_model = MultinomialNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)
print("Naive Bayes Classification Report:\n", classification_report(y_test, y_pred_nb))
```

Random Forest Model

```
rf_model = RandomForestClassifier(n_estimators=100, random_state = 42 )
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)
print("Random Forest Classification Report: \n", classification_report(y_test, y_pred_rf))
```

Logistic Regression Model

```
[27] model = LogisticRegression(max_iter=1000)
      model.fit(X_train, y_train)
```

LogisticRegression ⓘ ?

```
LogisticRegression(max_iter=1000)
```

```
y1= model.predict(X_train)
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
```

Model Validation and Evaluation Report (5 marks):

Model	Classification Report	F1 Score	Confusion Matrix																																						
KNN	<div>KNN Classification Report:<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>NEGATIVE</td><td>0.72</td><td>0.11</td><td>0.20</td><td>1386</td></tr><tr><td>NEUTRAL</td><td>0.25</td><td>0.87</td><td>0.38</td><td>1745</td></tr><tr><td>POSITIVE</td><td>0.96</td><td>0.63</td><td>0.76</td><td>9370</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.60</td><td>12501</td></tr><tr><td>macro avg</td><td>0.64</td><td>0.54</td><td>0.45</td><td>12501</td></tr><tr><td>weighted avg</td><td>0.83</td><td>0.60</td><td>0.64</td><td>12501</td></tr></tbody></table></div>		precision	recall	f1-score	support	NEGATIVE	0.72	0.11	0.20	1386	NEUTRAL	0.25	0.87	0.38	1745	POSITIVE	0.96	0.63	0.76	9370	accuracy			0.60	12501	macro avg	0.64	0.54	0.45	12501	weighted avg	0.83	0.60	0.64	12501	44.54%	<div>Confusion Matrix:<table><tbody><tr><td>[[158 1167 61]</td></tr><tr><td>[31 1512 202]</td></tr><tr><td>[30 3479 5861]]</td></tr></tbody></table></div>	[[158 1167 61]	[31 1512 202]	[30 3479 5861]]
	precision	recall	f1-score	support																																					
NEGATIVE	0.72	0.11	0.20	1386																																					
NEUTRAL	0.25	0.87	0.38	1745																																					
POSITIVE	0.96	0.63	0.76	9370																																					
accuracy			0.60	12501																																					
macro avg	0.64	0.54	0.45	12501																																					
weighted avg	0.83	0.60	0.64	12501																																					
[[158 1167 61]																																									
[31 1512 202]																																									
[30 3479 5861]]																																									
Naïve Bayes	<div>Naïve Bayes Classification Report:<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>NEGATIVE</td><td>0.75</td><td>0.76</td><td>0.76</td><td>1386</td></tr><tr><td>NEUTRAL</td><td>0.63</td><td>0.21</td><td>0.31</td><td>1745</td></tr><tr><td>POSITIVE</td><td>0.86</td><td>0.97</td><td>0.91</td><td>9370</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.84</td><td>12501</td></tr><tr><td>macro avg</td><td>0.75</td><td>0.65</td><td>0.66</td><td>12501</td></tr><tr><td>weighted avg</td><td>0.82</td><td>0.84</td><td>0.81</td><td>12501</td></tr></tbody></table></div>		precision	recall	f1-score	support	NEGATIVE	0.75	0.76	0.76	1386	NEUTRAL	0.63	0.21	0.31	1745	POSITIVE	0.86	0.97	0.91	9370	accuracy			0.84	12501	macro avg	0.75	0.65	0.66	12501	weighted avg	0.82	0.84	0.81	12501	66.11%	<div>Confusion Matrix:<table><tbody><tr><td>[[1054 98 234]</td></tr><tr><td>[193 365 1187]</td></tr><tr><td>[158 120 9092]]</td></tr></tbody></table></div>	[[1054 98 234]	[193 365 1187]	[158 120 9092]]
	precision	recall	f1-score	support																																					
NEGATIVE	0.75	0.76	0.76	1386																																					
NEUTRAL	0.63	0.21	0.31	1745																																					
POSITIVE	0.86	0.97	0.91	9370																																					
accuracy			0.84	12501																																					
macro avg	0.75	0.65	0.66	12501																																					
weighted avg	0.82	0.84	0.81	12501																																					
[[1054 98 234]																																									
[193 365 1187]																																									
[158 120 9092]]																																									

Random Forest	<table><tr><th colspan="6">Random Forest Classification Report:</th></tr><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th><th></th></tr><tr><td>NEGATIVE</td><td>0.83</td><td>0.61</td><td>0.70</td><td>1386</td><td></td></tr><tr><td>NEUTRAL</td><td>0.79</td><td>0.58</td><td>0.67</td><td>1745</td><td></td></tr><tr><td>POSITIVE</td><td>0.90</td><td>0.97</td><td>0.93</td><td>9370</td><td></td></tr><tr><td>accuracy</td><td></td><td></td><td>0.88</td><td>12501</td><td></td></tr><tr><td>macro avg</td><td>0.84</td><td>0.72</td><td>0.77</td><td>12501</td><td></td></tr><tr><td>weighted avg</td><td>0.87</td><td>0.88</td><td>0.87</td><td>12501</td><td></td></tr></table>	Random Forest Classification Report:							precision	recall	f1-score	support		NEGATIVE	0.83	0.61	0.70	1386		NEUTRAL	0.79	0.58	0.67	1745		POSITIVE	0.90	0.97	0.93	9370		accuracy			0.88	12501		macro avg	0.84	0.72	0.77	12501		weighted avg	0.87	0.88	0.87	12501		76.92	<table><tr><th colspan="4">Confusion Matrix:</th></tr><tr><td>[[</td><td>849</td><td>98</td><td>439]</td></tr><tr><td>[</td><td>97</td><td>1019</td><td>629]</td></tr><tr><td>[</td><td>83</td><td>170</td><td>9117]]</td></tr></table>	Confusion Matrix:				[[849	98	439]	[97	1019	629]	[83	170	9117]]
Random Forest Classification Report:																																																																			
	precision	recall	f1-score	support																																																															
NEGATIVE	0.83	0.61	0.70	1386																																																															
NEUTRAL	0.79	0.58	0.67	1745																																																															
POSITIVE	0.90	0.97	0.93	9370																																																															
accuracy			0.88	12501																																																															
macro avg	0.84	0.72	0.77	12501																																																															
weighted avg	0.87	0.88	0.87	12501																																																															
Confusion Matrix:																																																																			
[[849	98	439]																																																																
[97	1019	629]																																																																
[83	170	9117]]																																																																
Logistic Regression	<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th><th></th></tr><tr><td>NEGATIVE</td><td>0.84</td><td>0.74</td><td>0.79</td><td>1386</td><td></td></tr><tr><td>NEUTRAL</td><td>0.74</td><td>0.66</td><td>0.70</td><td>1745</td><td></td></tr><tr><td>POSITIVE</td><td>0.93</td><td>0.97</td><td>0.95</td><td>9370</td><td></td></tr><tr><td>accuracy</td><td></td><td></td><td>0.90</td><td>12501</td><td></td></tr><tr><td>macro avg</td><td>0.84</td><td>0.79</td><td>0.81</td><td>12501</td><td></td></tr><tr><td>weighted avg</td><td>0.89</td><td>0.90</td><td>0.90</td><td>12501</td><td></td></tr></table>		precision	recall	f1-score	support		NEGATIVE	0.84	0.74	0.79	1386		NEUTRAL	0.74	0.66	0.70	1745		POSITIVE	0.93	0.97	0.95	9370		accuracy			0.90	12501		macro avg	0.84	0.79	0.81	12501		weighted avg	0.89	0.90	0.90	12501		81.1%	<table><tr><th colspan="4">Confusion Matrix:</th></tr><tr><td>[[1021</td><td>180</td><td>185]</td></tr><tr><td>[114</td><td>1155</td><td>476]</td></tr><tr><td>[78</td><td>229</td><td>9063]]</td></tr></table>	Confusion Matrix:				[[1021	180	185]	[114	1155	476]	[78	229	9063]]									
	precision	recall	f1-score	support																																																															
NEGATIVE	0.84	0.74	0.79	1386																																																															
NEUTRAL	0.74	0.66	0.70	1745																																																															
POSITIVE	0.93	0.97	0.95	9370																																																															
accuracy			0.90	12501																																																															
macro avg	0.84	0.79	0.81	12501																																																															
weighted avg	0.89	0.90	0.90	12501																																																															
Confusion Matrix:																																																																			
[[1021	180	185]																																																																	
[114	1155	476]																																																																	
[78	229	9063]]																																																																	