Analiza vectorizer (1, 2) min_df
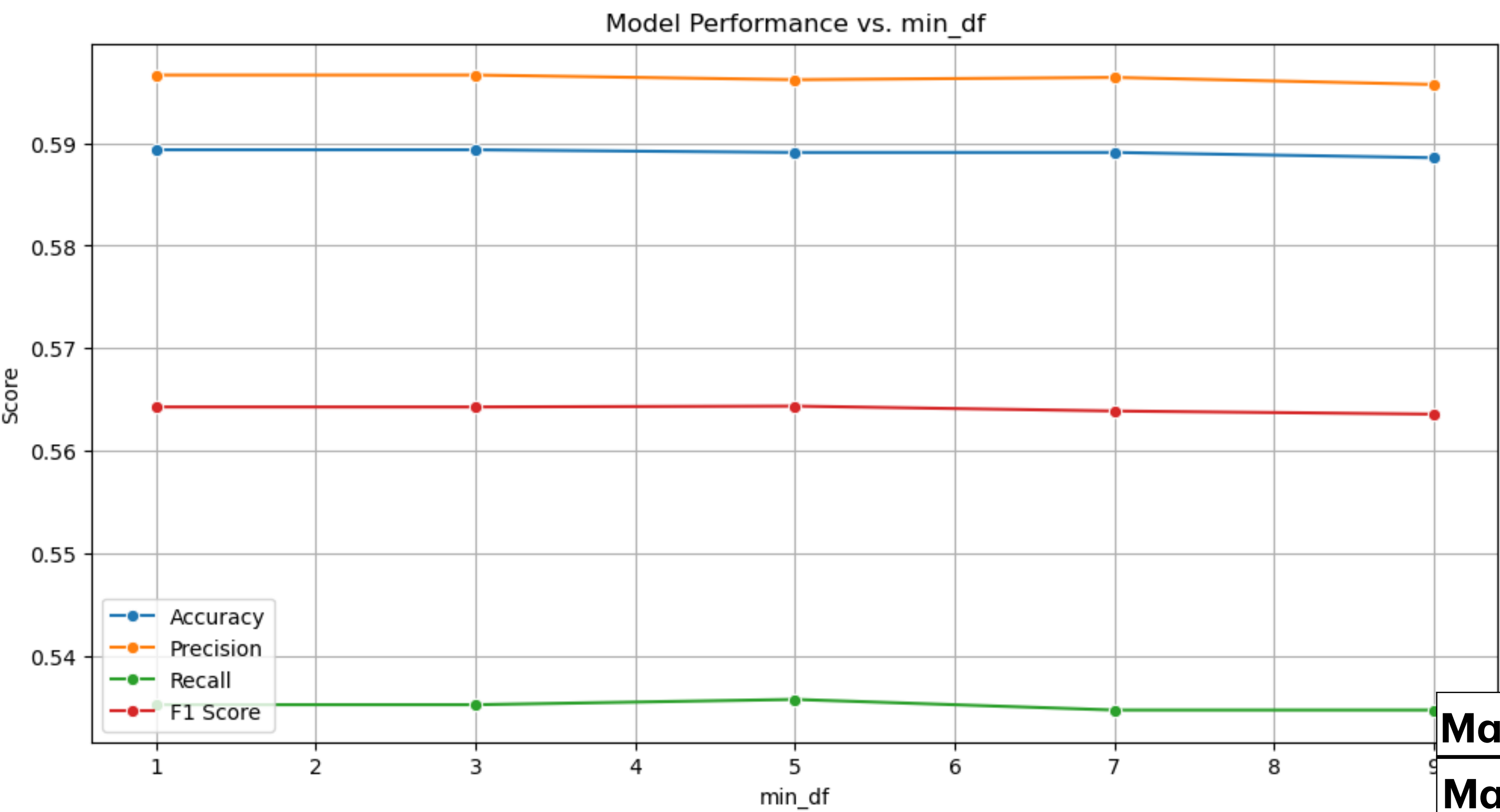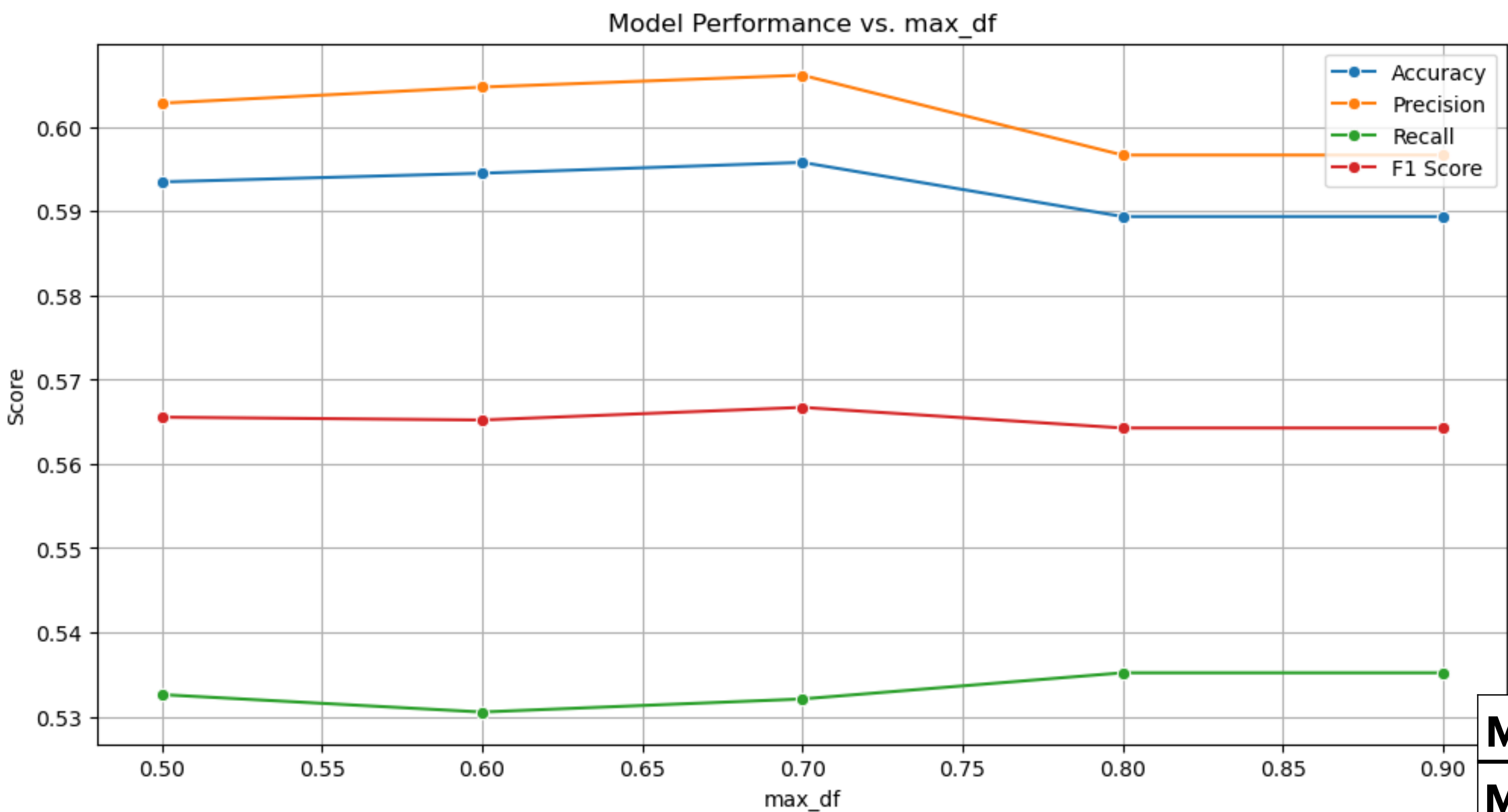
min_df= range(1, 10, 2)
 TfidfVectorizer(ngram_range=(1, 2), max_features=5000, min_df=min_d, max_df=0.9)

Max accuracy: 0.589	**min_df: 1.0**	**max_df: 0.9**
Max precision: 0.597	min_df: 1.0	max_df: 0.9
Max recall: 0.536	min_df: 5.0	max_df: 0.9
Max f1: 0.564	min_df: 5.0	max_df: 0.9

Analiza vectorizer (1, 2) max_df

max_df= [0.5, 0.6, 0.7, 0.8, 0.9], min_df = 1
vectorizer = TfidfVectorizer(ngram_range=(1, 2), max_features=5000, min_df=min_df, max_df=max_d)

Model Performance vs. max_df

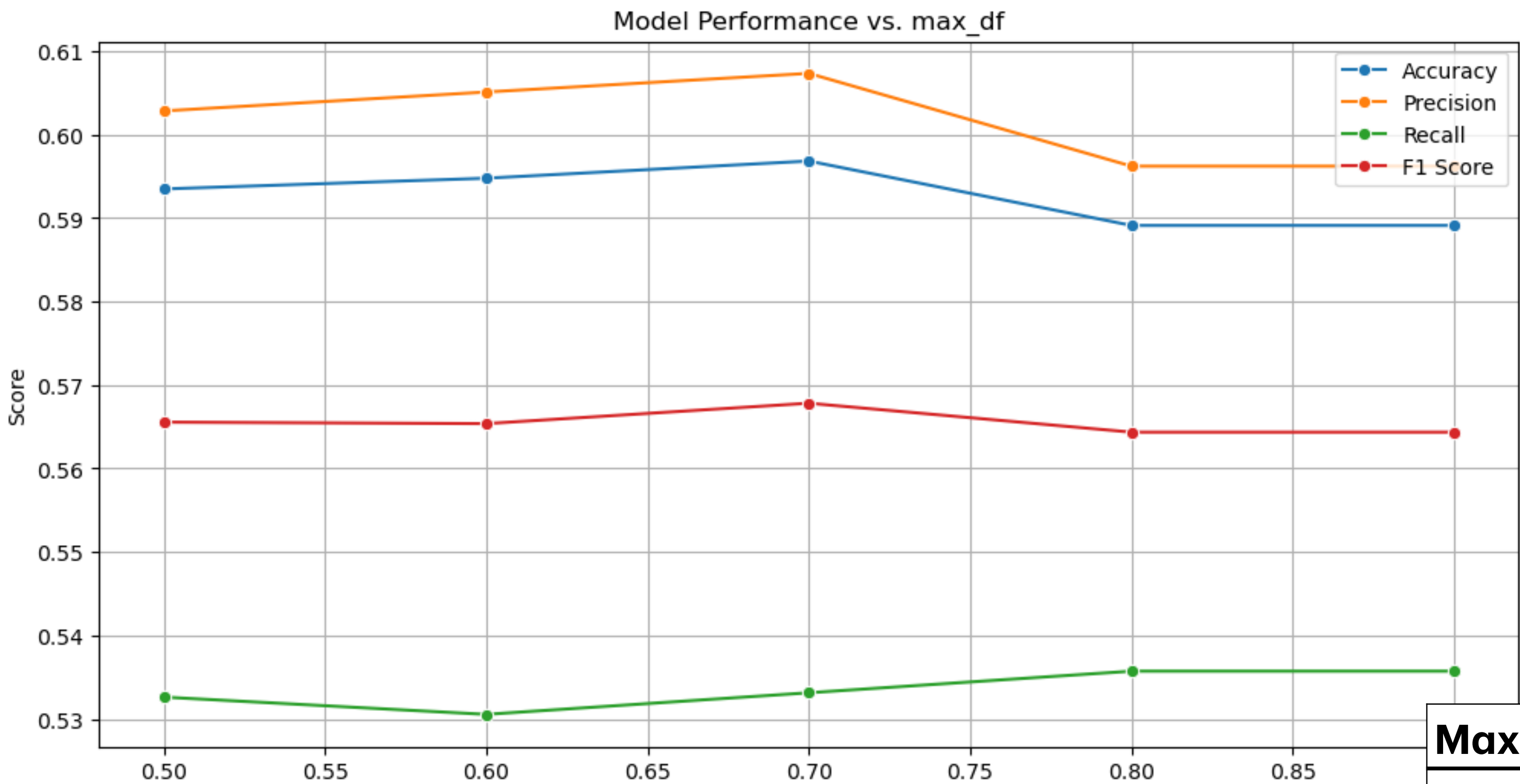| Max accuracy: 0.596	max_df: 0.7	min_df: 1.0
Max precision: 0.606	max_df: 0.7	min_df: 1.0
Max recall: 0.535	max_df: 0.8	min_df: 1.0
Max f1: 0.567	max_df: 0.7	min_df: 1.0

Analiza vectorizer (1, 2) max_df

max_df= [0.5, 0.6, 0.7, 0.8, 0.9], min_df = 5
vectorizer = TfidfVectorizer(ngram_range=(1, 2), max_features=5000, min_df=min_df, max_df=max_d)

vectorizer = TfidfVectorizer(ngram_range=(1, 2),
    max_features=5000,
    min_df=5,
    max_df=0.7)

Max accuracy: 0.597	max_df: 0.7	min_df: 5.0
Max precision: 0.607	max_df: 0.7	min_df: 5.0
Max recall: 0.536	max_df: 0.8	min_df: 5.0
Max f1: 0.568	max_df: 0.7	min_df: 5.0

Analiza vectorizer (1, 3) min_df

min_df= range(1, 10, 2)
Vectorizer = TfidfVectorizer(ngram_range=(1, 3), max_features=5000, min_df=min_d, max_df=0.9)

Model Performance vs. min_df

Max accuracy: 0.595	min_df: 7.0	max_df: 0.9
Max precision: 0.605	min_df: 7.0	max_df: 0.9
Max recall: 0.536	min_df: 9.0	max_df: 0.9
Max f1: 0.568	min_df: 9.0	max_df: 0.9

Analiza vectorizer (1, 3) max_df

max_df= [0.5, 0.6, 0.7, 0.8, 0.9], min_df = 7
vectorizer = TfidfVectorizer(ngram_range=(1, 3), max_features=5000, min_df=min_df, max_df=max_d)

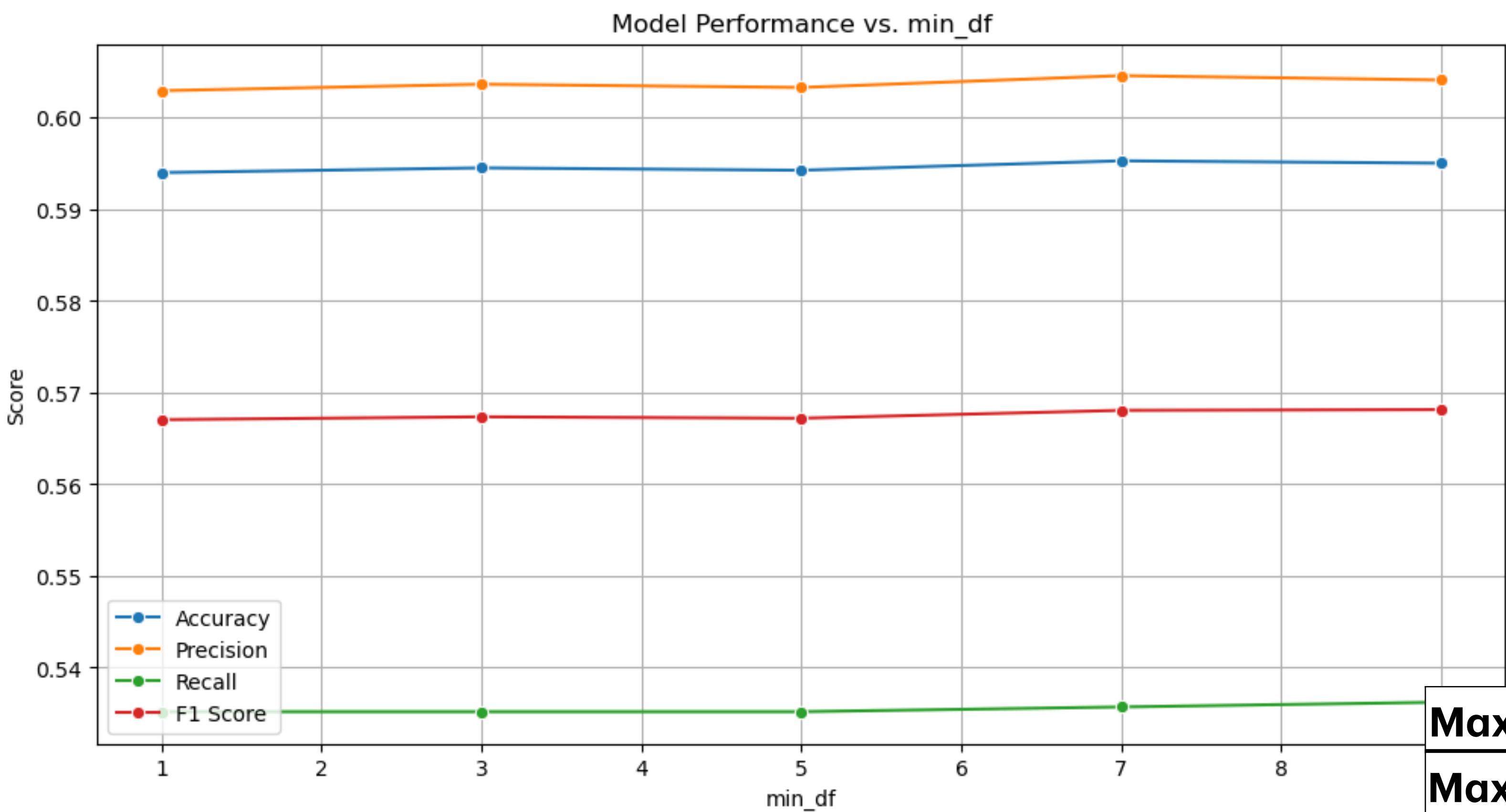Model Performance vs. max_df
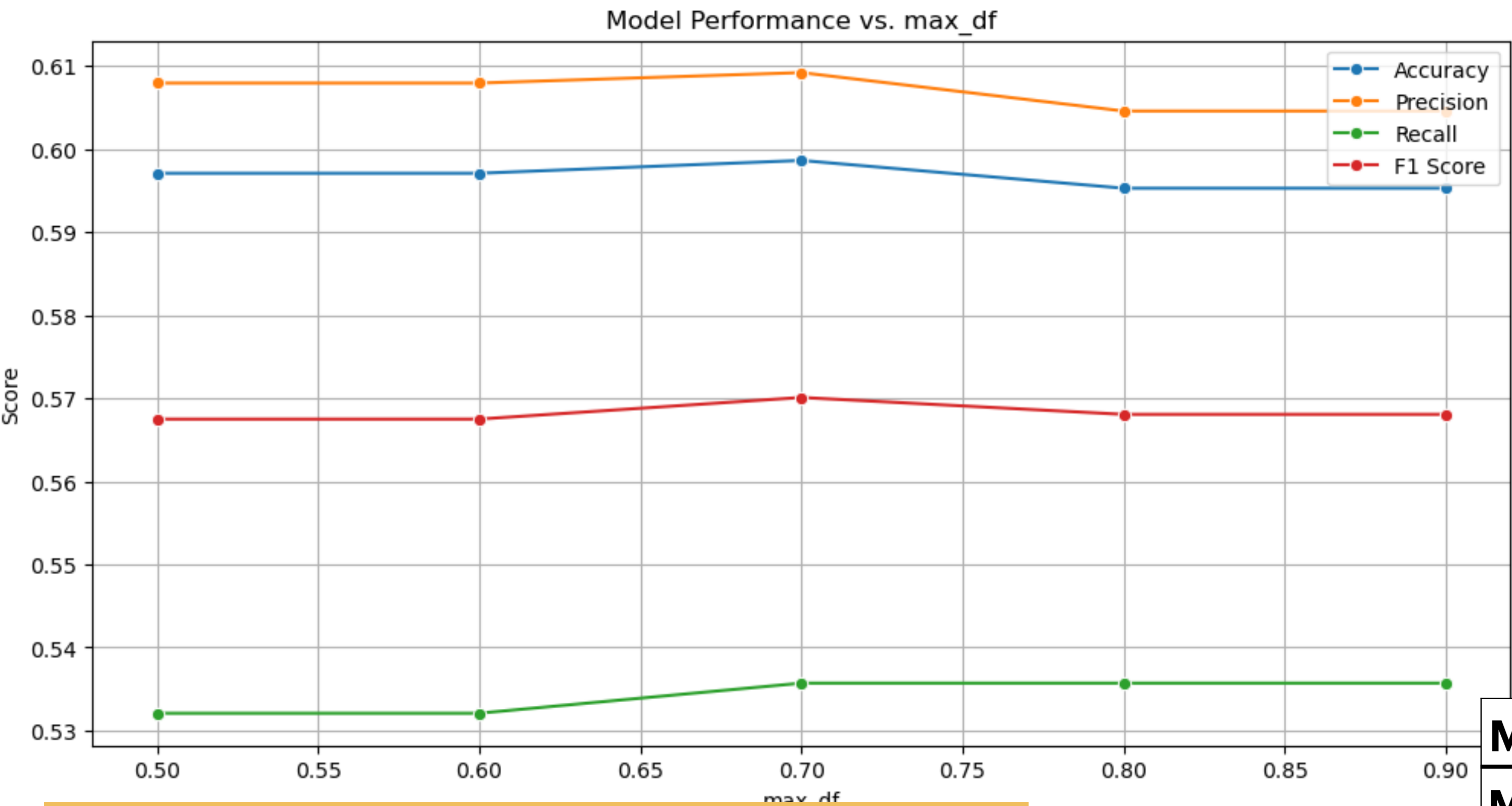
vectorizer = TfidfVectorizer(ngram_range=(1, 3),
    max_features=5000,
    min_df=7,
    max_df=0.7)

| Max accuracy: 0.599	min_df: 7.0	max_df: 0.7
Max precision: 0.609	min_df: 7.0	max_df: 0.7
Max recall: 0.536	min_df: 7.0	max_df: 0.7
Max f1: 0.570	min_df: 7.0	max_df: 0.7

Logistic regression
GridSearch

```
Najlepsze parametry:
{'C': 1, 'class_weight': 'balanced', 'penalty': 'l2', 'solver': 'liblinear'}

Najlepszy wynik F1-score (na walidacji krzyżowej): 0.5678
Accuracy: 0.5793
Precision: 0.5824

Ocena na zbiorze testowym:
              precision    recall   f1-score    support

           0       0.58      0.60       0.59       1948
           1       0.58      0.56       0.57       1941

    accuracy                            0.58       3889
   macro avg       0.58      0.58       0.58       3889
weighted avg       0.58      0.58       0.58       3889
```


Confusion Matrix for Logistic Regression
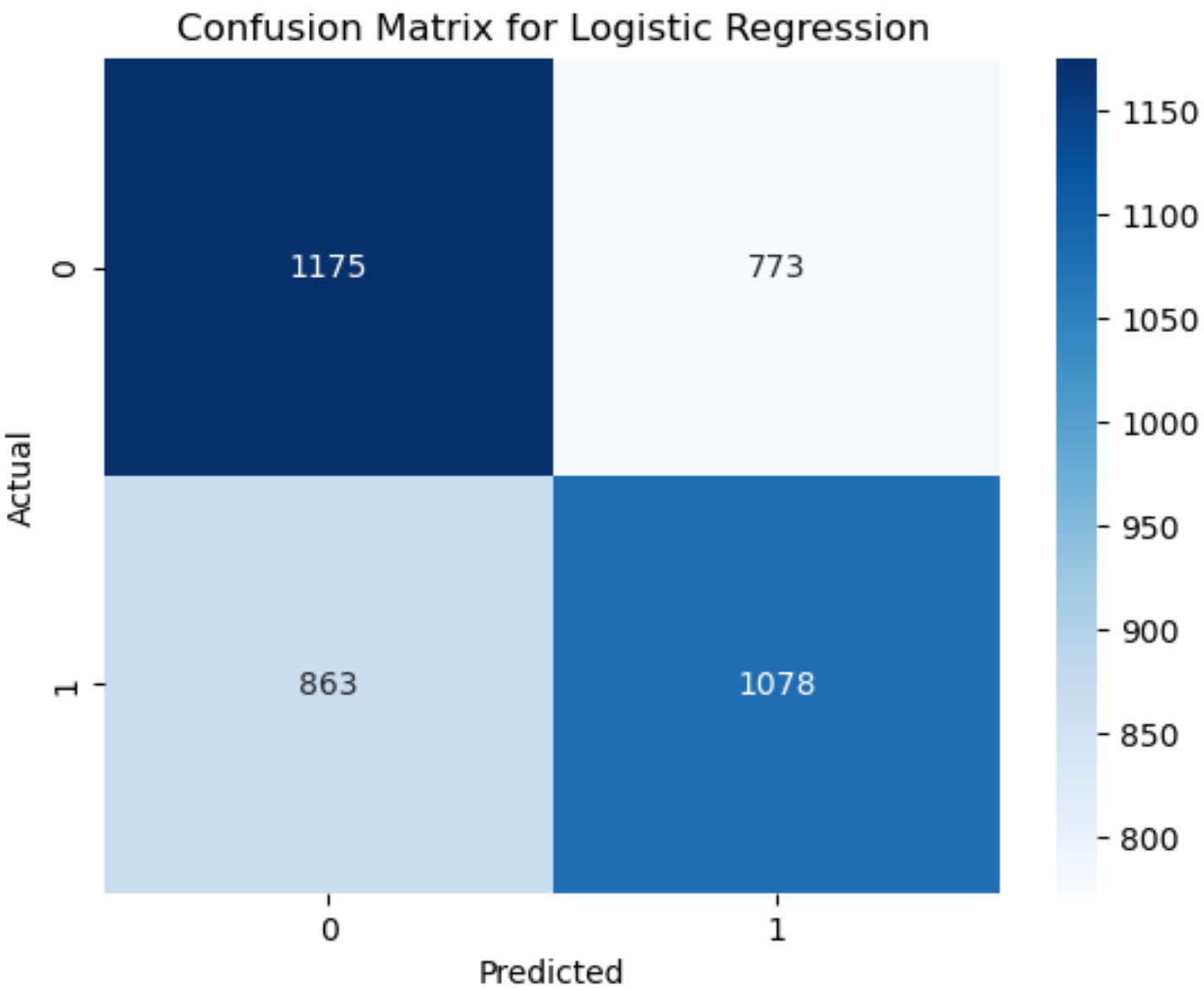
LGBM

GridSearch

Najlepsze parametry: {'learning_rate': 0.1, 'n_estimators': 100, 'num_leaves': 31}
Najlepszy F1-score (walidacja krzyżowa): 0.5635
Accuracy: 0.5649
Precision: 0.5680

Raport klasyfikacji na zbiorze testowym:

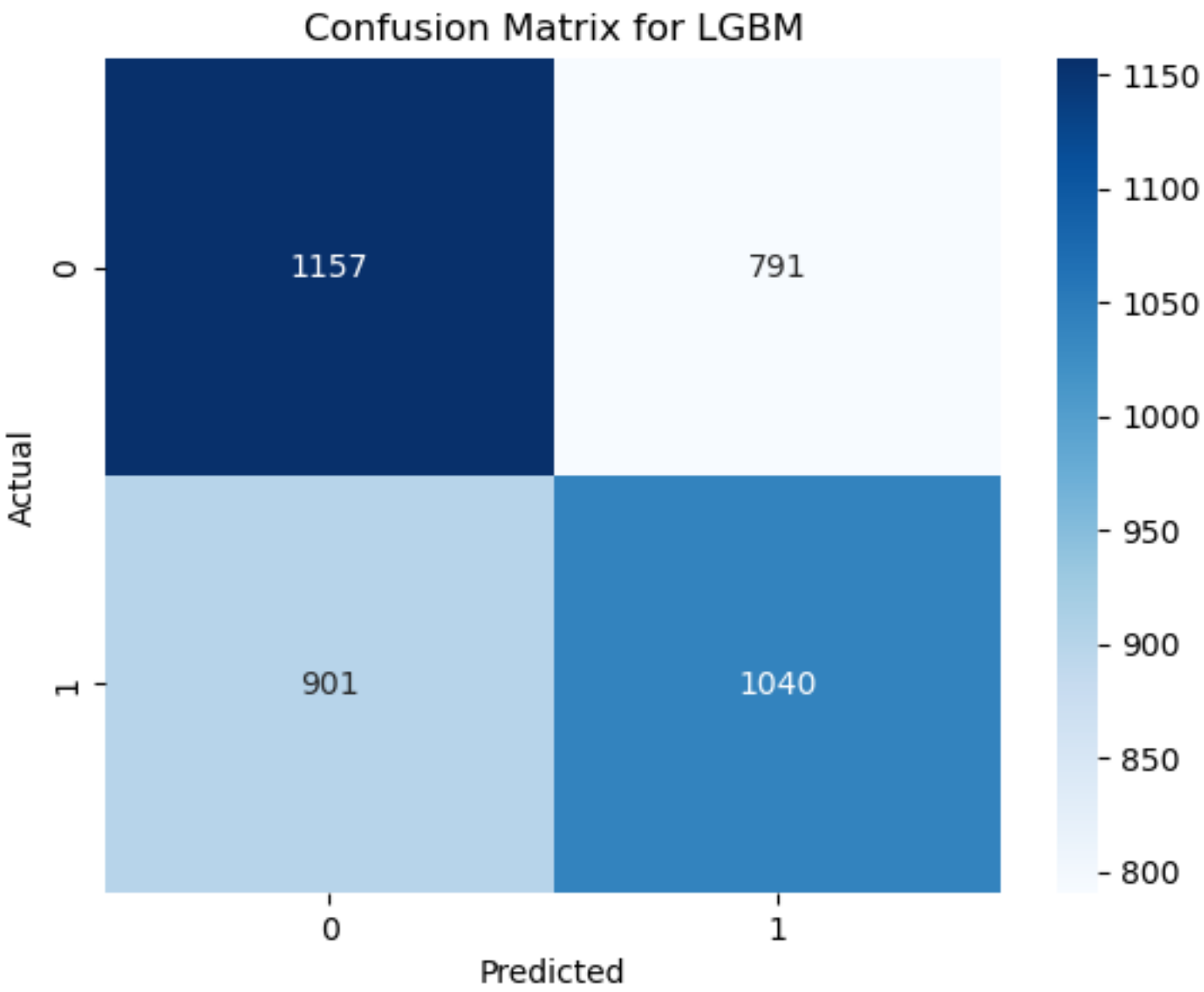|	precision	recall	f1-score	support
0	0.56	0.59	0.58	1948
1	0.57	0.54	0.55	1941
accuracy			0.56	3889
macro avg	0.57	0.56	0.56	3889
weighted avg	0.57	0.56	0.56	3889

Confusion Matrix for LGBM
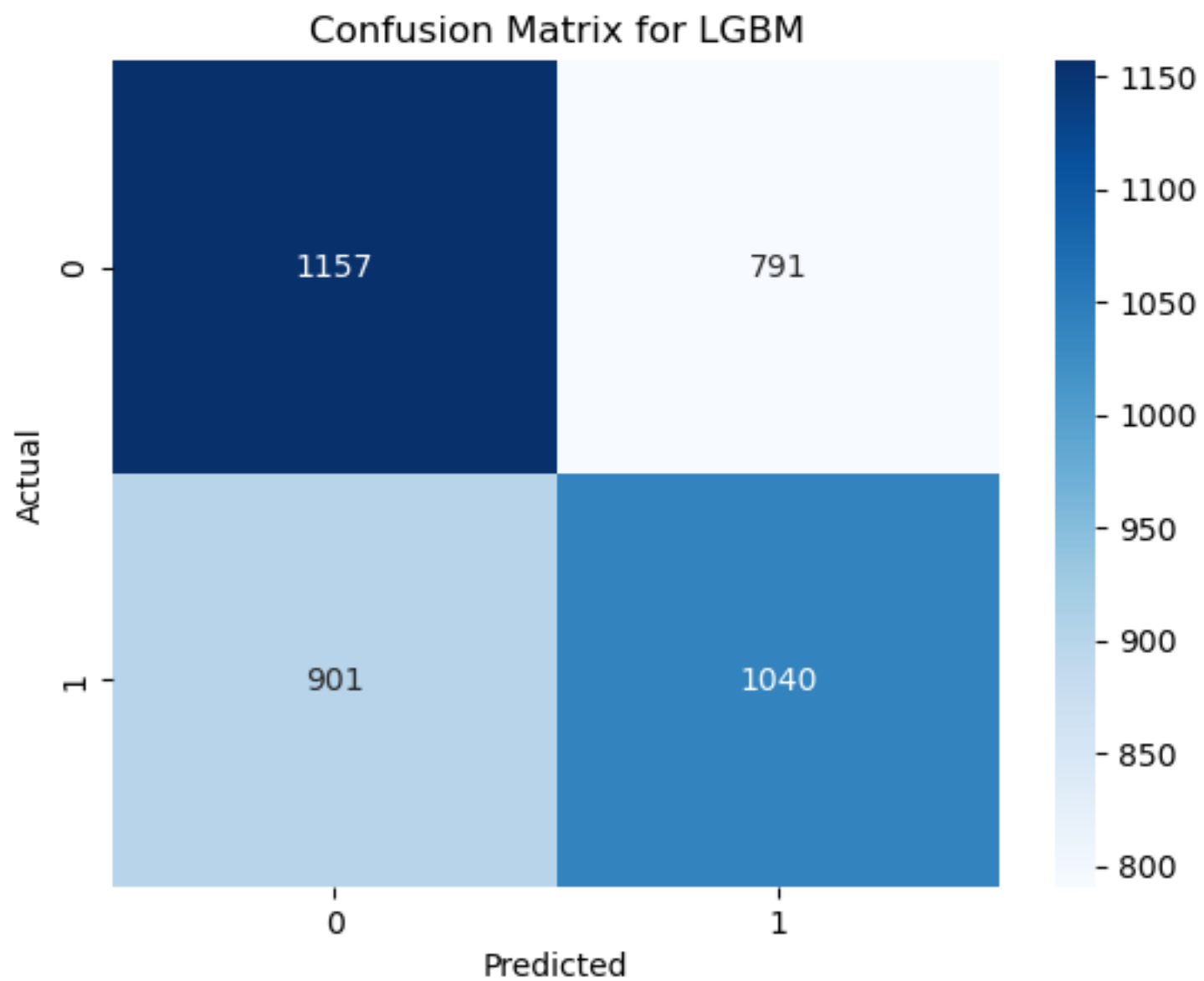
XGBoost

GridSearch

Najlepsze parametry: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 300}
Najlepszy F1-score (walidacja krzyżowa): 0.5529
Accuracy: 0.5930
Precision: 0.6084

Raport klasyfikacji na zbiorze testowym:

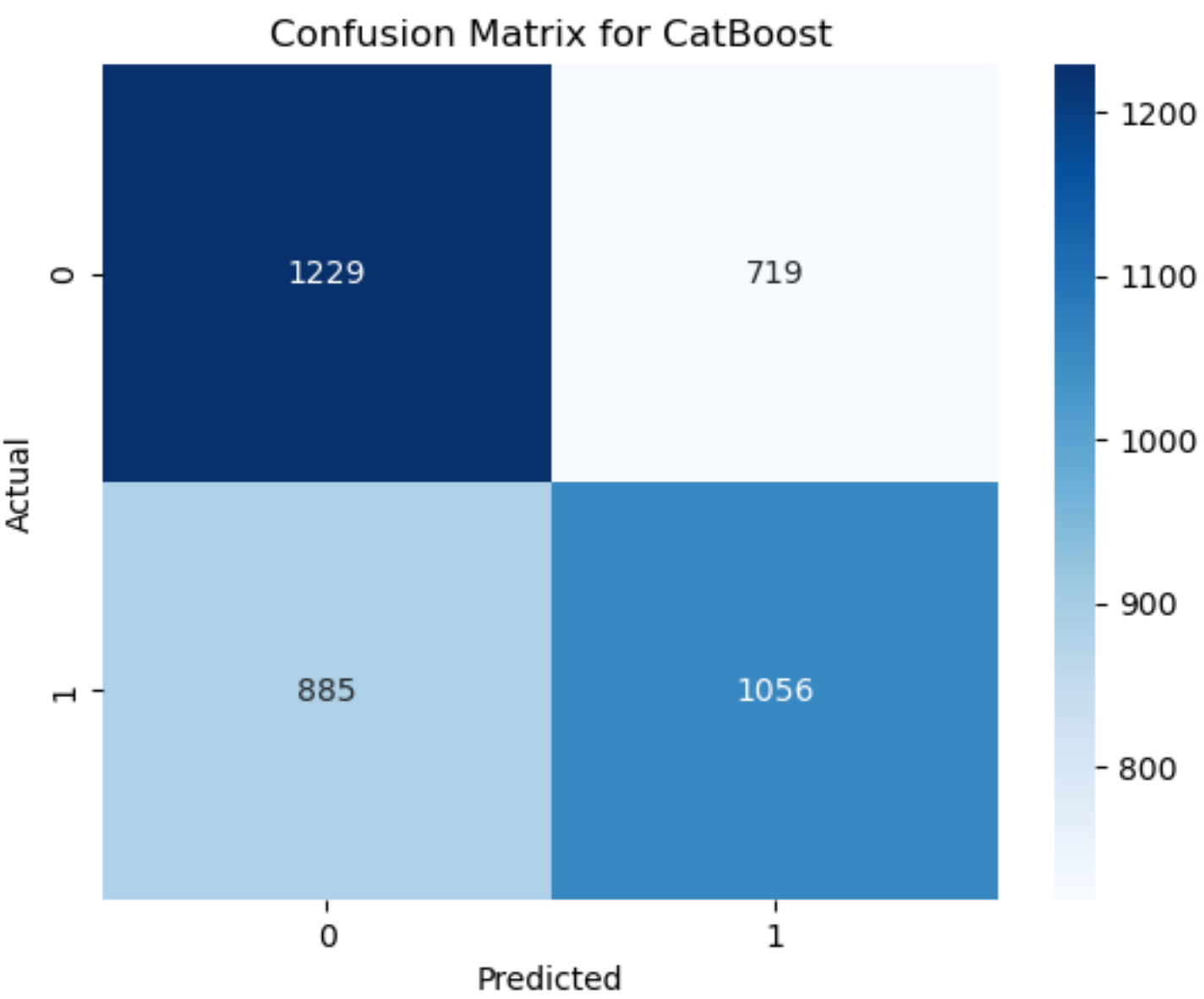|	precision	recall	f1-score	support
0	0.58	0.67	0.62	1948
1	0.61	0.52	0.56	1941
accuracy			0.59	3889
macro avg	0.59	0.59	0.59	3889
weighted avg	0.59	0.59	0.59	3889

Confusion Matrix for LGBM

CatBoost

GridSearch

Najlepsze parametry: {'depth': 3, 'iterations': 500, 'learning_rate': 0.1}
Najlepszy F1-score (walidacja krzyżowa): 0.5628
Accuracy: 0.5876
Precision: 0.5949

Raport klasyfikacji na zbiorze testowym:

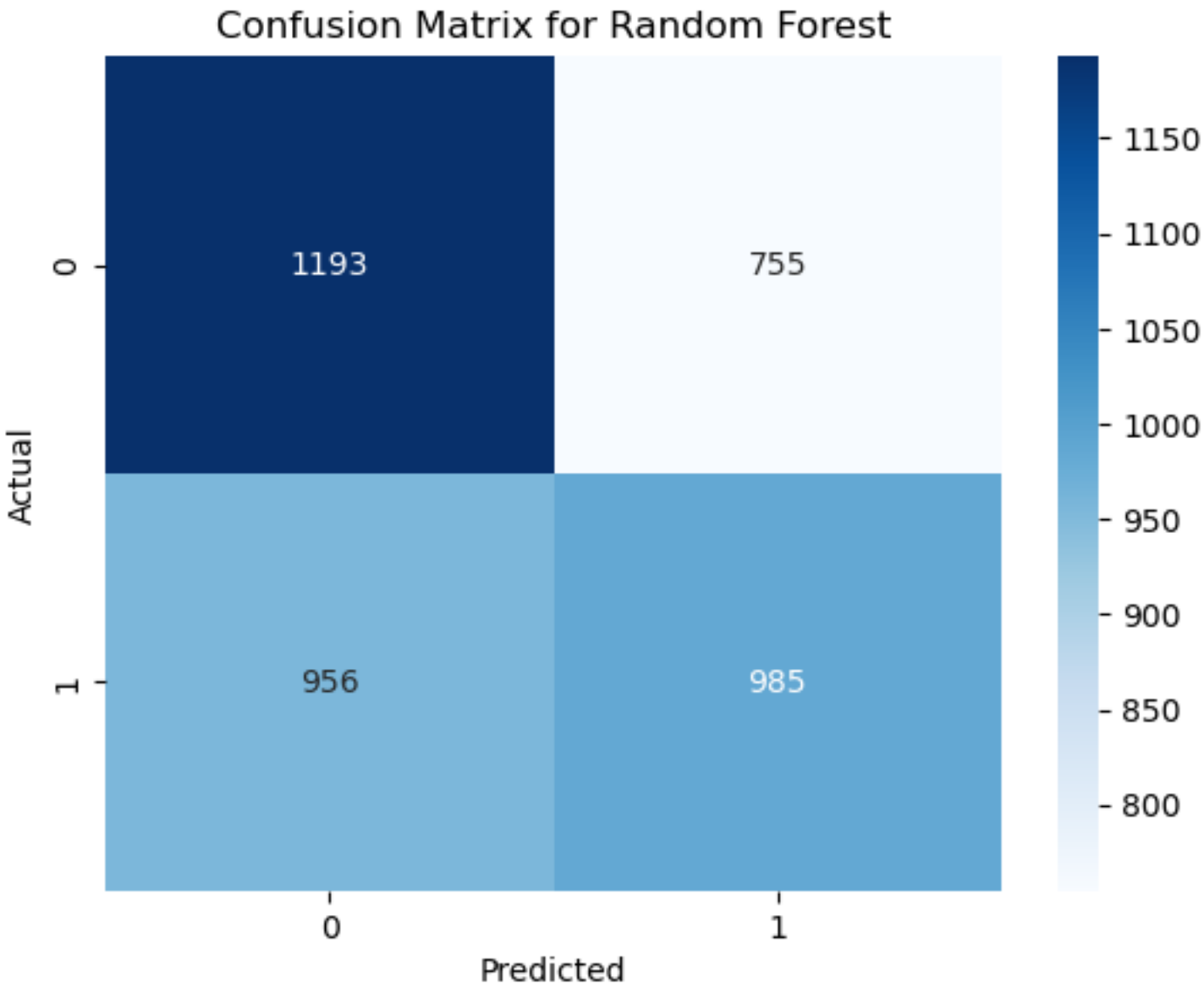|	precision	recall	f1-score	support
0	0.58	0.63	0.61	1948
1	0.59	0.54	0.57	1941
accuracy			0.59	3889
macro avg	0.59	0.59	0.59	3889
weighted avg	0.59	0.59	0.59	3889

Confusion Matrix for CatBoost

Random Forest

GridSearch

```
Najlepsze parametry: {'max_depth': None, 'min_samples_split': 2, 'n_estimators': 100}
Najlepszy F1-score (walidacja krzyżowa): 0.5415
Accuracy: 0.5600
Precision: 0.5661
```

```
Raport klasyfikacji na zbiorze testowym:
              precision    recall  f1-score   support

           0       0.56      0.61      0.58      1948
           1       0.57      0.51      0.54      1941

    accuracy                           0.56      3889
   macro avg       0.56      0.56      0.56      3889
weighted avg       0.56      0.56      0.56      3889
```
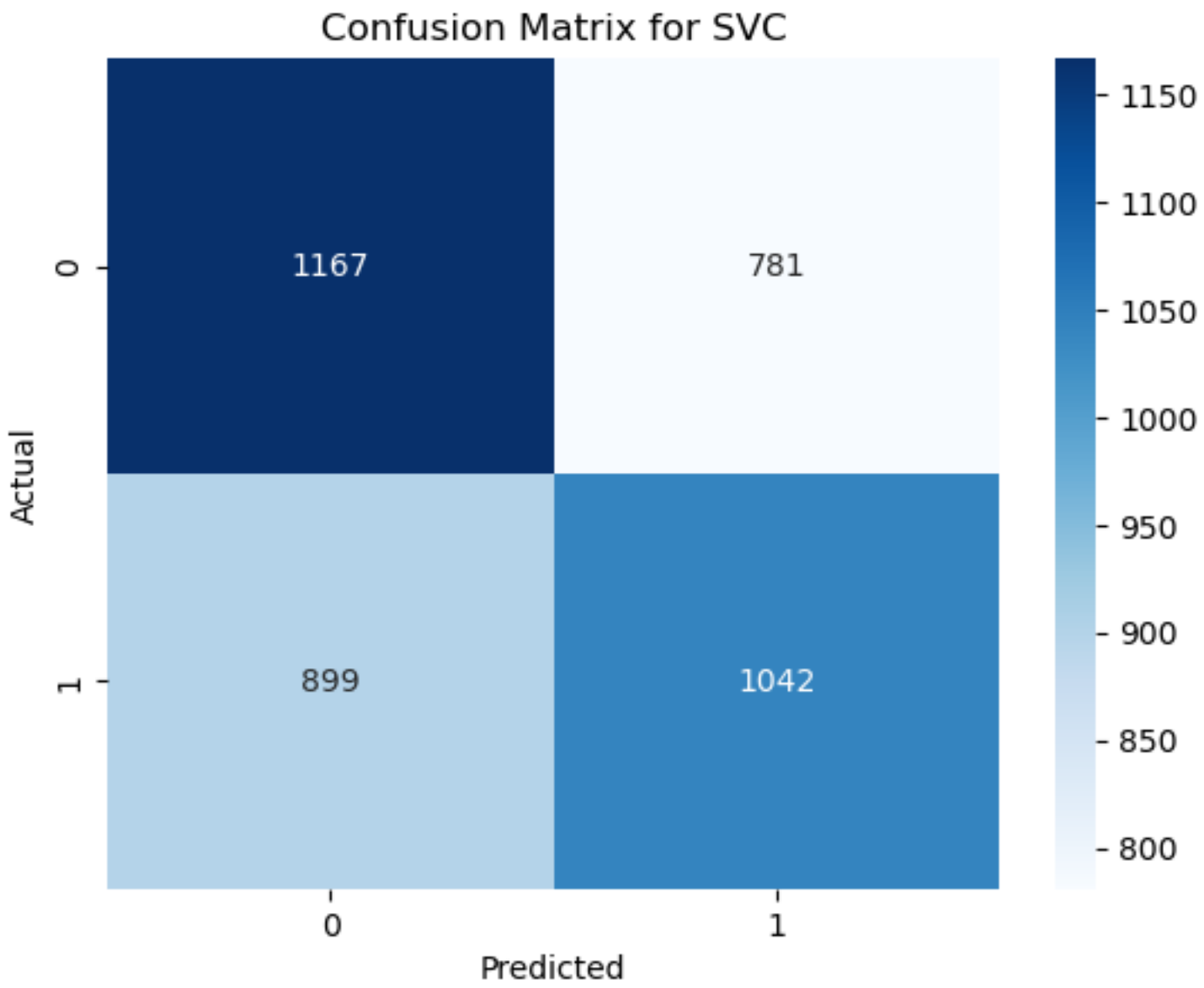

Confusion Matrix for Random Forest

SVC

GridSearch small

```
Najlepsze parametry: {'C': 1, 'kernel': 'linear'}
Najlepszy F1-score (walidacja krzyżowa): 0.5551
Accuracy: 0.5680
Precision: 0.5716

Raport klasyfikacji na zbiorze testowym:
              precision    recall  f1-score   support

           0       0.56      0.60      0.58      1948
           1       0.57      0.54      0.55      1941

    accuracy                           0.57      3889
   macro avg       0.57      0.57      0.57      3889
weighted avg       0.57      0.57      0.57      3889
```


Confusion Matrix for SVC

SVC

GridSearch advanced + rbf kernel

```
Najlepsze parametry: {'C': 1, 'class_weight': 'balanced', 'kernel': 'linear'}
Najlepszy F1-score (walidacja krzyżowa): 0.5561
Accuracy: 0.5680
Precision: 0.5714

Raport klasyfikacji na zbiorze testowym:
              precision    recall   f1-score    support

           0      0.57      0.60       0.58       1948
           1      0.57      0.54       0.55       1941

    accuracy                           0.57       3889
   macro avg      0.57      0.57       0.57       3889
weighted avg      0.57      0.57       0.57       3889
```
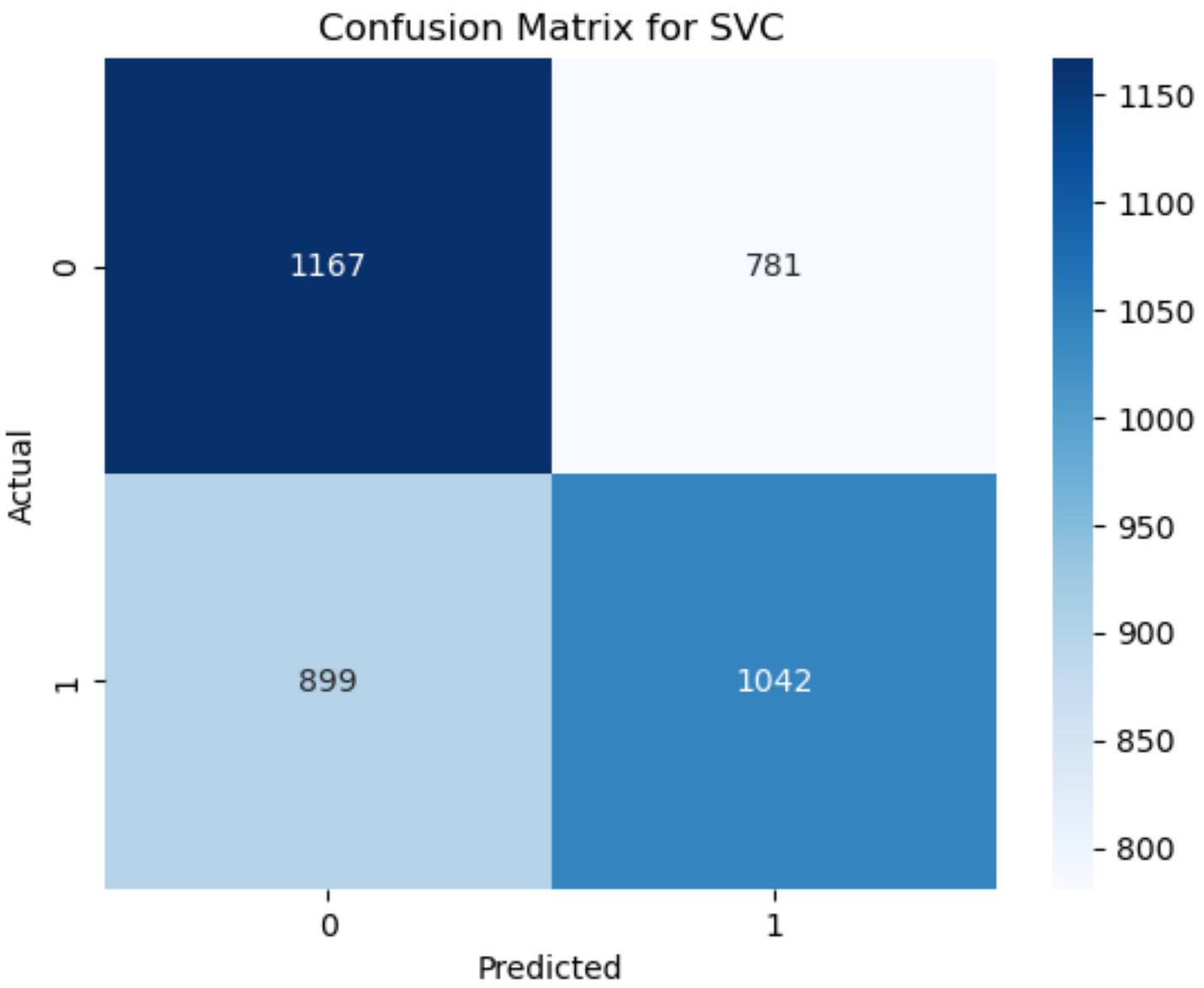

Confusion Matrix for SVC

SVC

GridSearch poly

Najlepsze parametry: {'C': 10, 'coef0': 0, 'degree': 2, 'gamma': 'scale', 'kernel': 'poly'}
Najlepszy F1-score (walidacja krzyżowa): 0.5482
Accuracy: 0.5492
Precision: 0.5489

Raport klasyfikacji na zbiorze testowym:

	precision	recall	f1-score	support
0	0.55	0.55	0.55	1948
1	0.55	0.54	0.55	1941
accuracy			0.55	3889
macro avg	0.55	0.55	0.55	3889
weighted avg	0.55	0.55	0.55	3889

Confusion Matrix for SVC Poly