Wine Quality Analytics System for BlueBerry Winery

Analytics & Research, Michal Pasternak

Agenda

1. Model
    a. Goal
    b. Selection process
    c. Optimization
    d. Limitations
    e. Results
    f. Final thoughts
2. Recommendations

Goal

Create a model that can detect the quality of wine.

Selection process

Testing algorithms which are used for Multi-Class Classification:

- K-Nearest Neighbours
- Decision trees
- Gradient Boosting
- Random Forest
- AdaBoost

Selection process

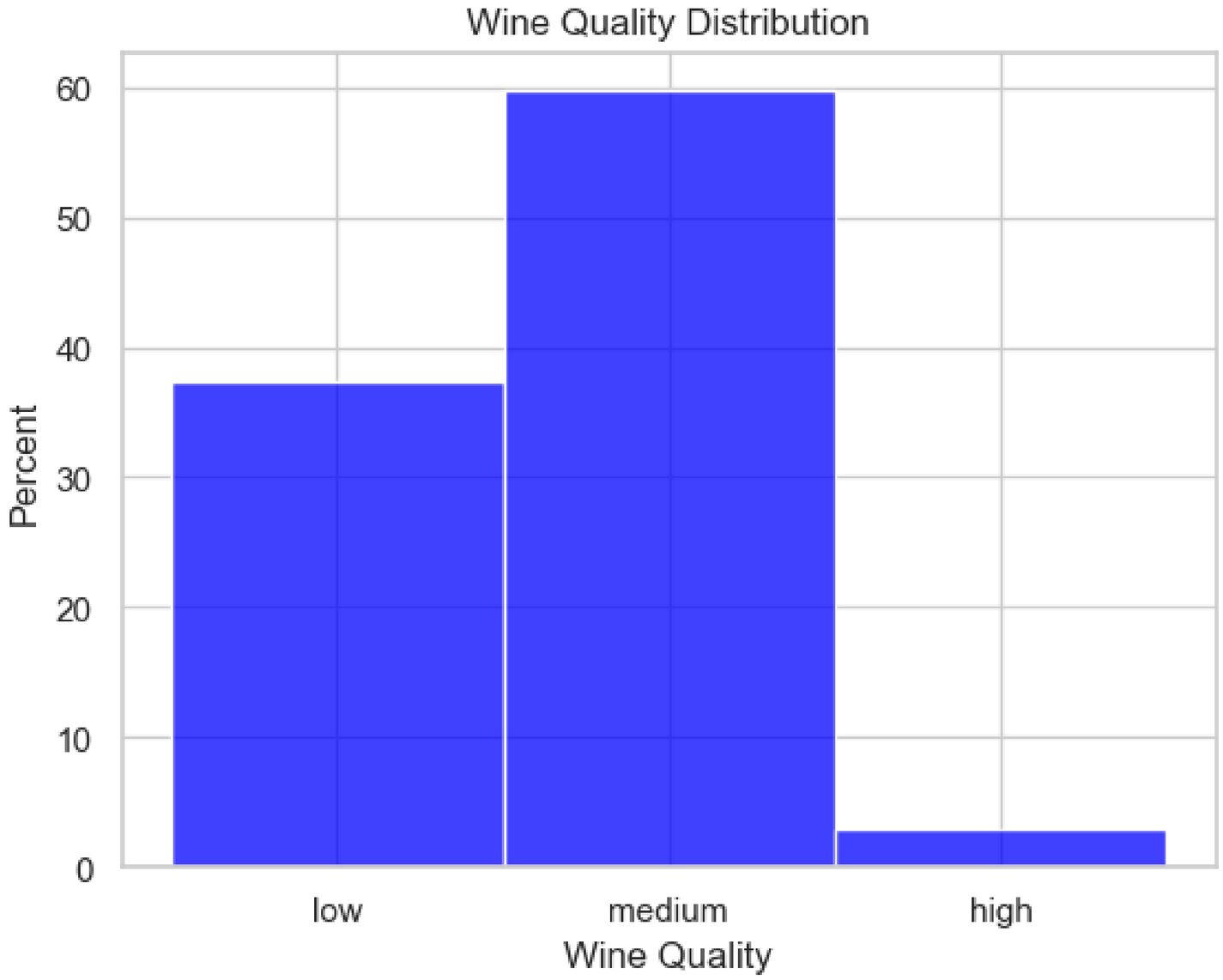
	model	accuracy	precision	recall	f1	kappa
0	KNeighborsClassifier()	0.640605	0.406980	0.413537	0.409309	0.227653
1	DecisionTreeClassifier()	0.633039	0.437433	0.437891	0.437027	0.269964
2	GradientBoostingClassifier()	0.722573	0.474545	0.472700	0.471703	0.402389
3	RandomForestClassifier()	0.738966	0.484194	0.490374	0.486740	0.444871
4	AdaBoostClassifier()	0.663304	0.478610	0.454672	0.457758	0.301139

Optimization

Methods tried to optimize the model:

- Dimension Reduction
- Hyperparameter Tuning
- Undersampling

Limitations

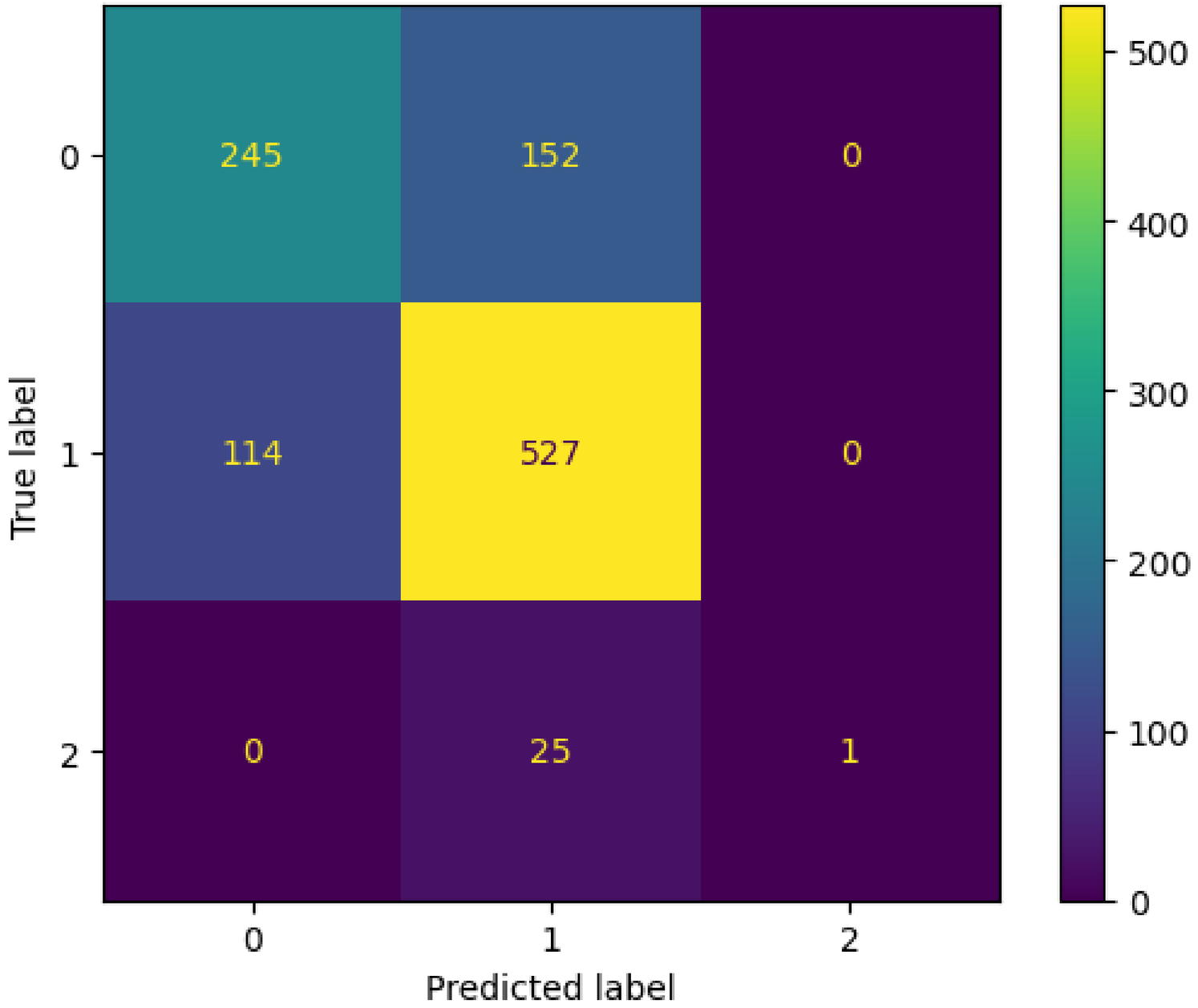


Results

accuracy	precision	recall	f1	kappa
0.726504	0.730049	0.726504	0.715748	0.424793

	precision	recall	f1-score	support
0	0.68	0.62	0.65	397
1	0.75	0.82	0.78	641
2	1.00	0.04	0.07	26
accuracy			0.73	1064
macro avg	0.81	0.49	0.50	1064
weighted avg	0.73	0.73	0.72	1064

Results


245
152
0
114
527
0
0
25
1
0
1
2
True label
0
1
2
Predicted label
500
400
300
200
100
0

Final thoughts

- The model could be used to roughly classify the wine
- The model should improve with more data
- Quality is subjective

Recommendations

- As per the EDA presented last time, there is a trend showing what chemical composition high quality wine should have and we shoukld try to produce more of it based on that
- Keep using the service of wine experts to expand the database
- Use the model to preclassify the data to limit spending on wine experts

Thanks you!