Automating Workflows for the City of Montgomery

BRANNON WALDEN

Contents

1. Introduction
2. Explanation of Data
3. Feature Engineering
4. Model Analysis
5. Model Summary
6. Next Steps
7. Appendix

Introduction

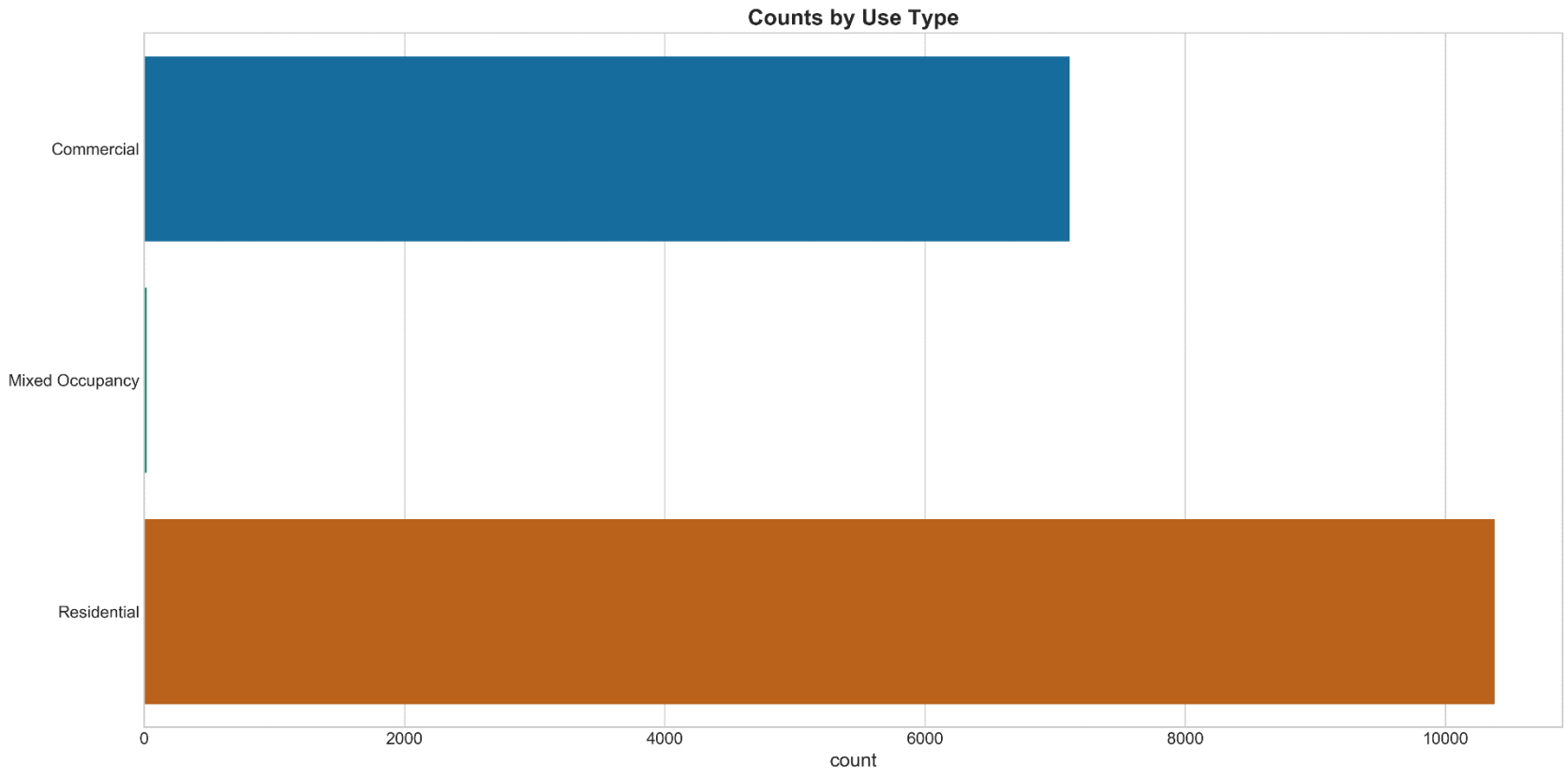
The City of Montgomery Inspections Department:

- Issues building permits for construction projects.
- Documents permit applications by entering long descriptions of the projects scope along with many other fields such as location, owner, estimated costs, permit types, zoning, etc.
- Data entry is time consuming and diverts staff attention away from code enforcement and other tasks.

Hypothesis:

The Use Type field, and possibly Job Type, could be automagically populated without human input and with better than average (>52.81%) results using natural language processing and machine learning techniques.

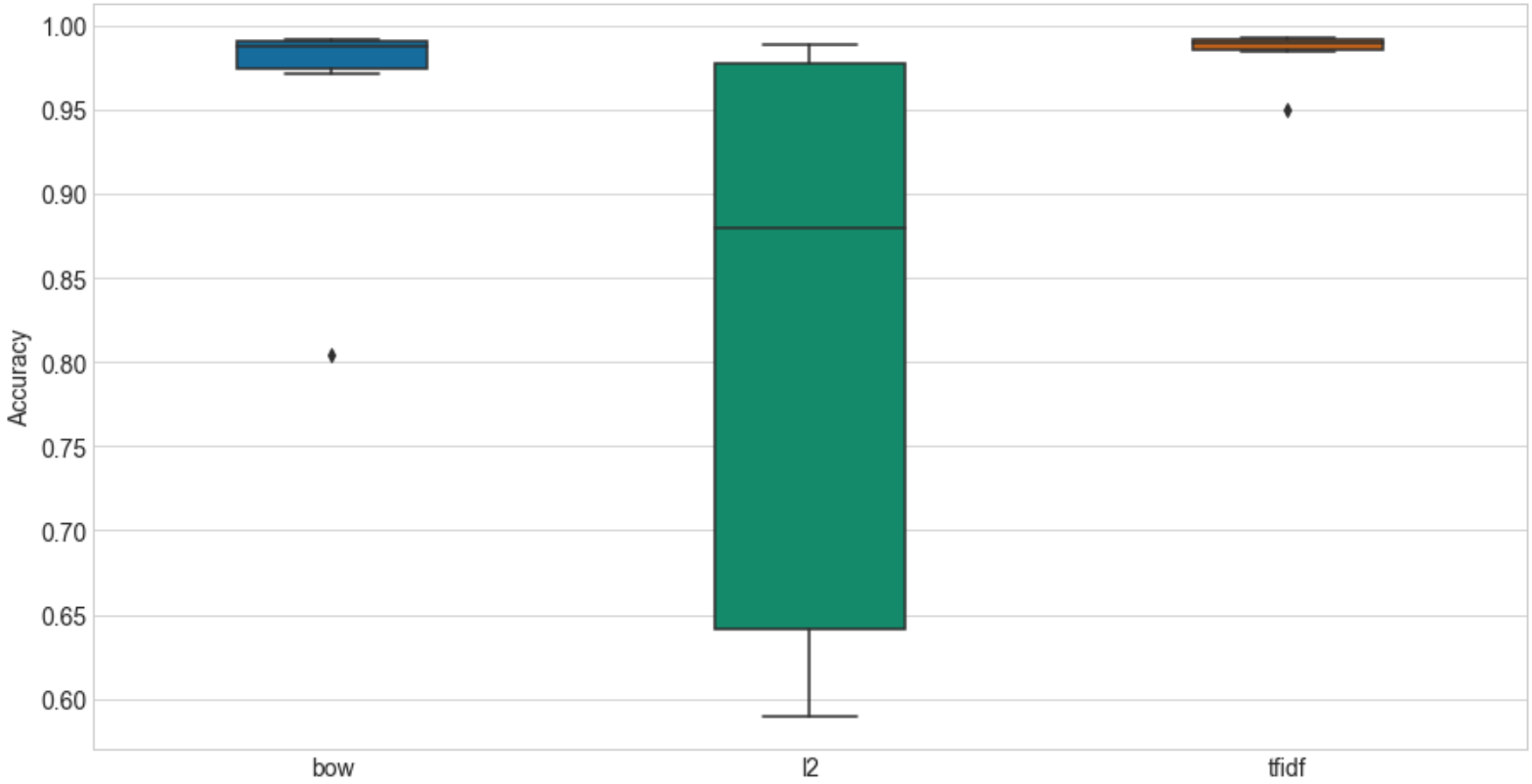
Explanation of Data



After removing Mixed Occupancy values, which are rare occurrences, the Use Type field looks good for binary classification.

Baseline metric = 52.81% accuracy

Feature Engineering

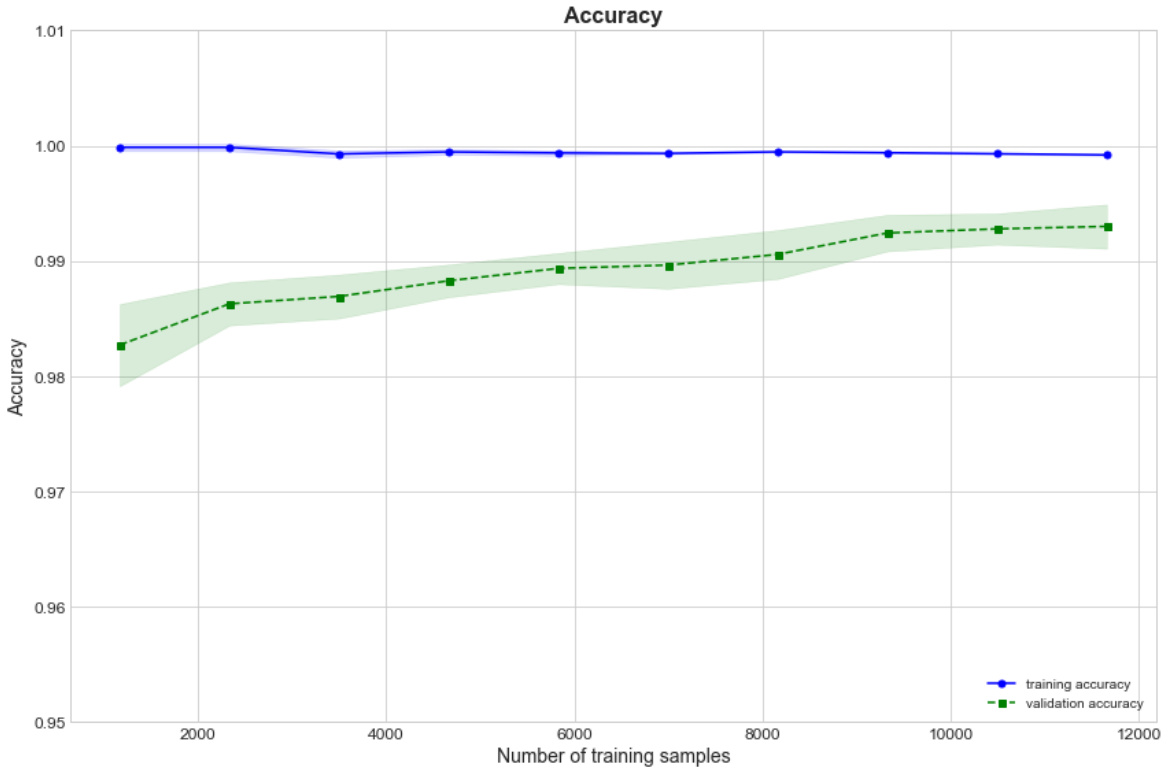


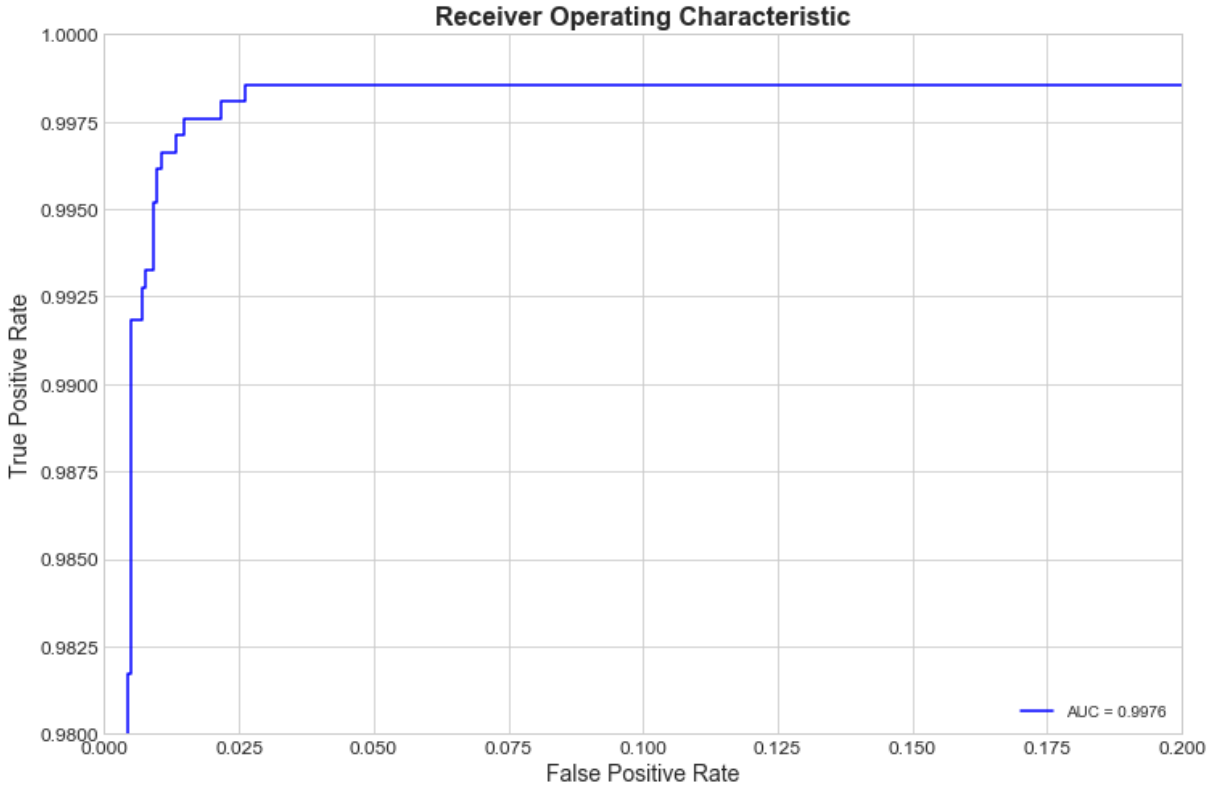
The Description Field was converted to Bag of Words, l2 normalized and tf-idf representations and compared using logistic regression.

The results indicated the tf-idf representation as the top performer.

Model Analysis

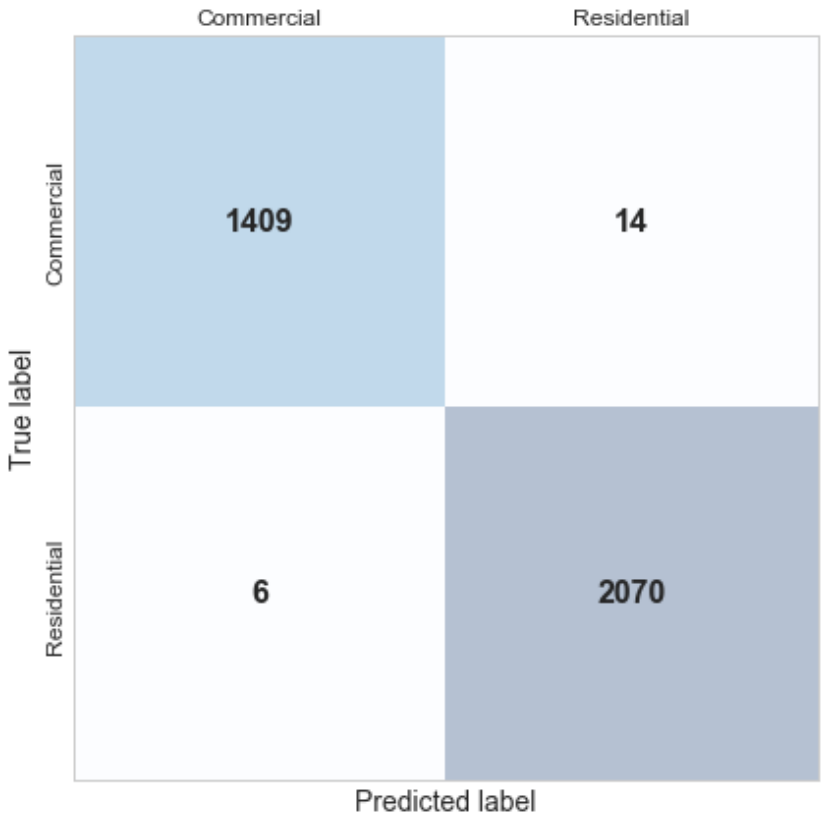
- Logistic Regression

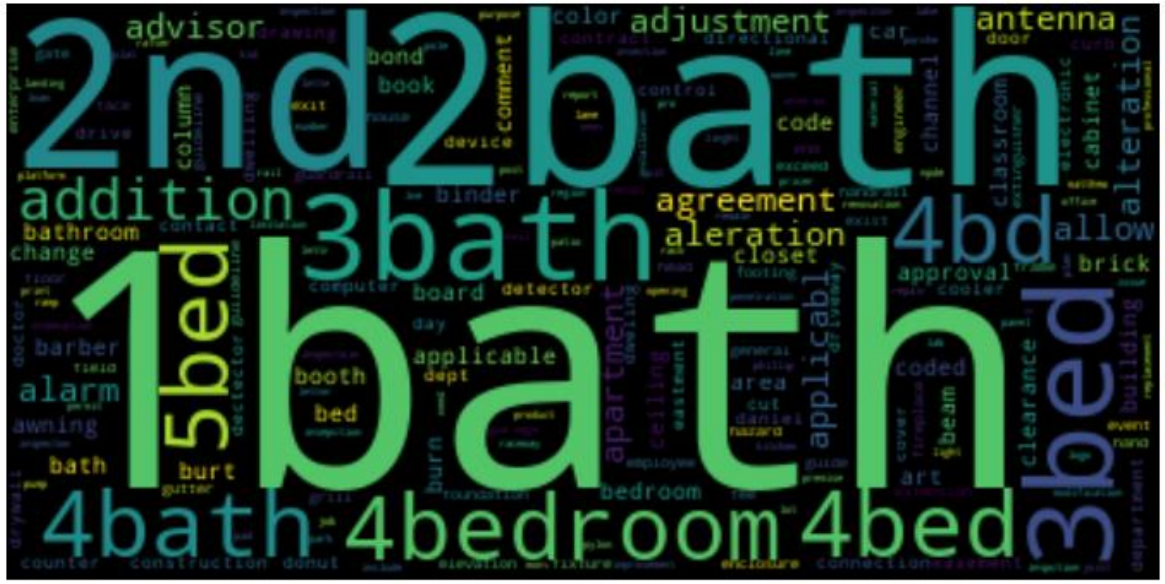




Model Analysis

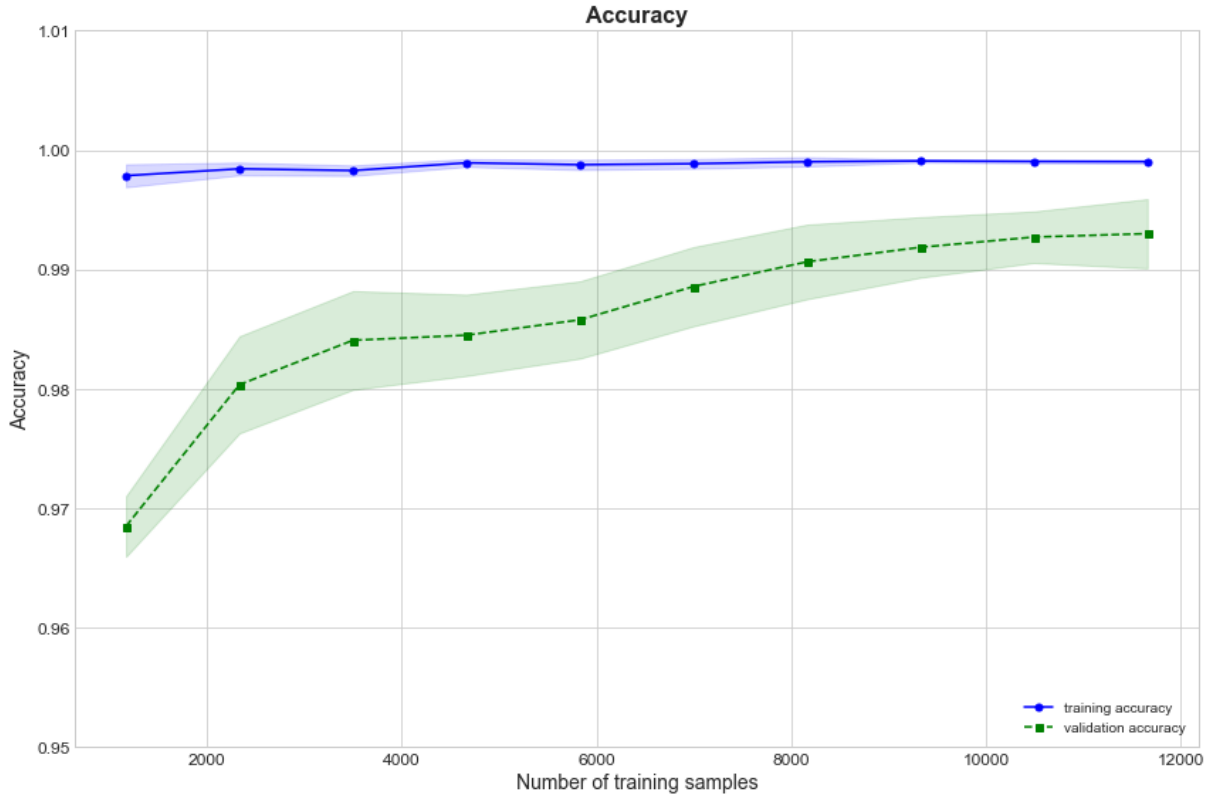
- Logistic Regression

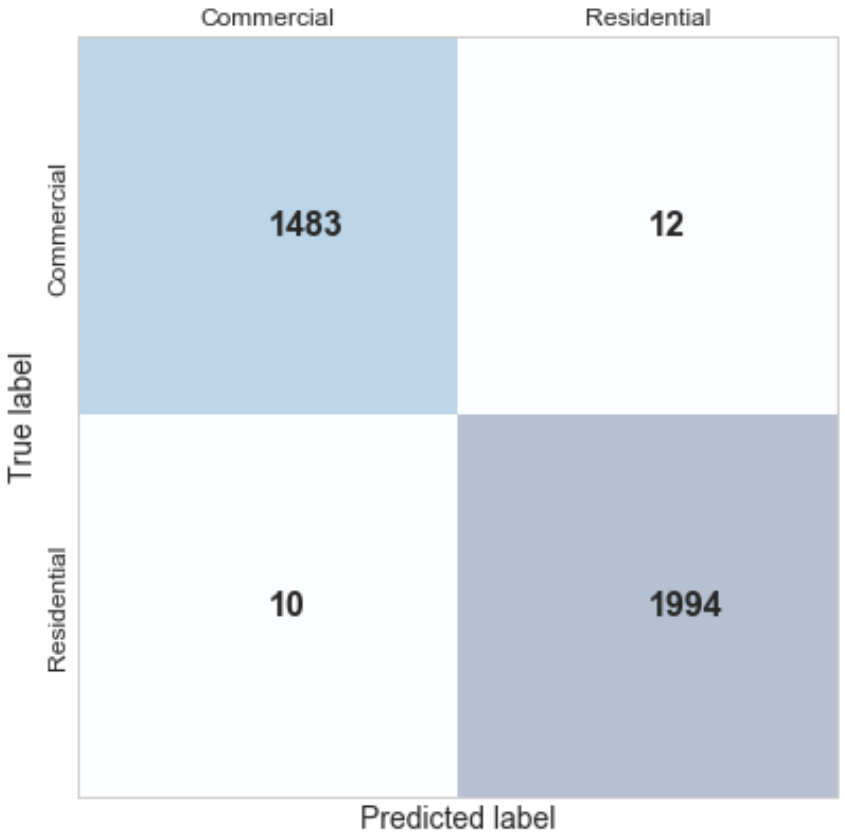




Model Analysis

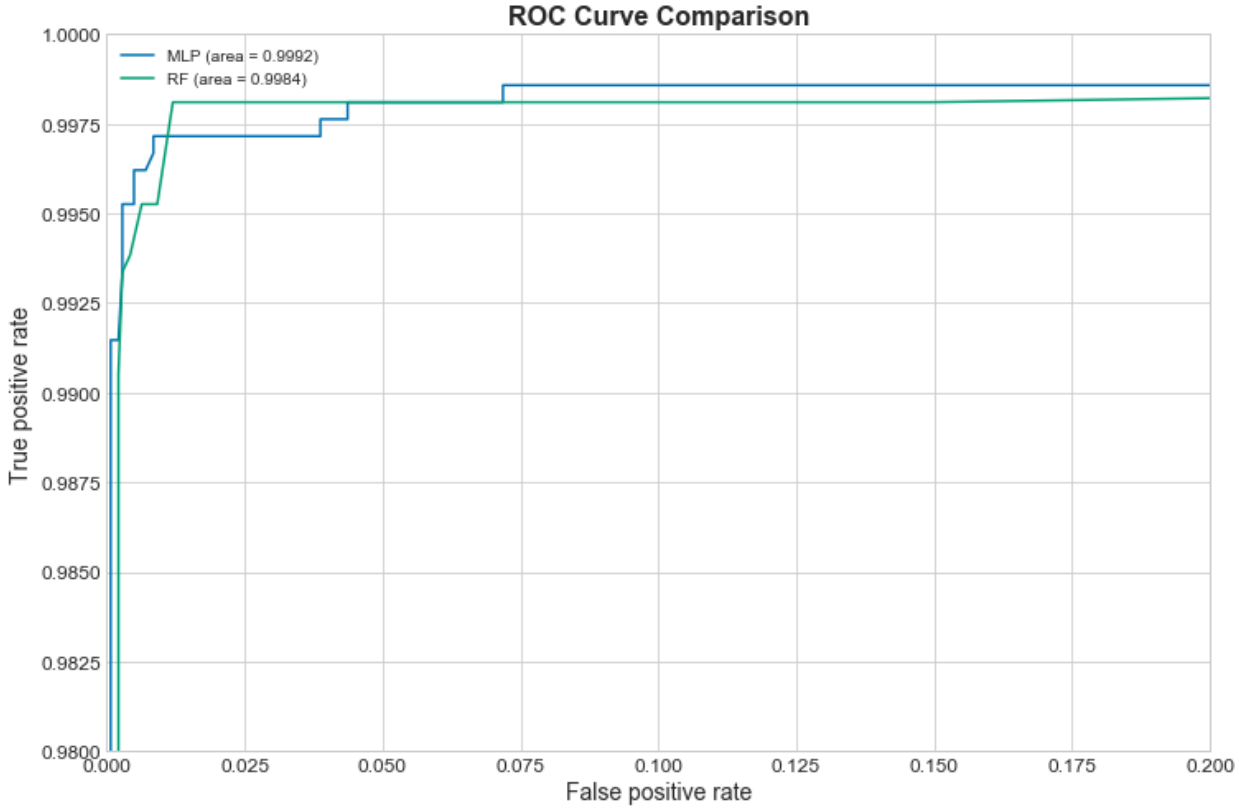
- Random Forest

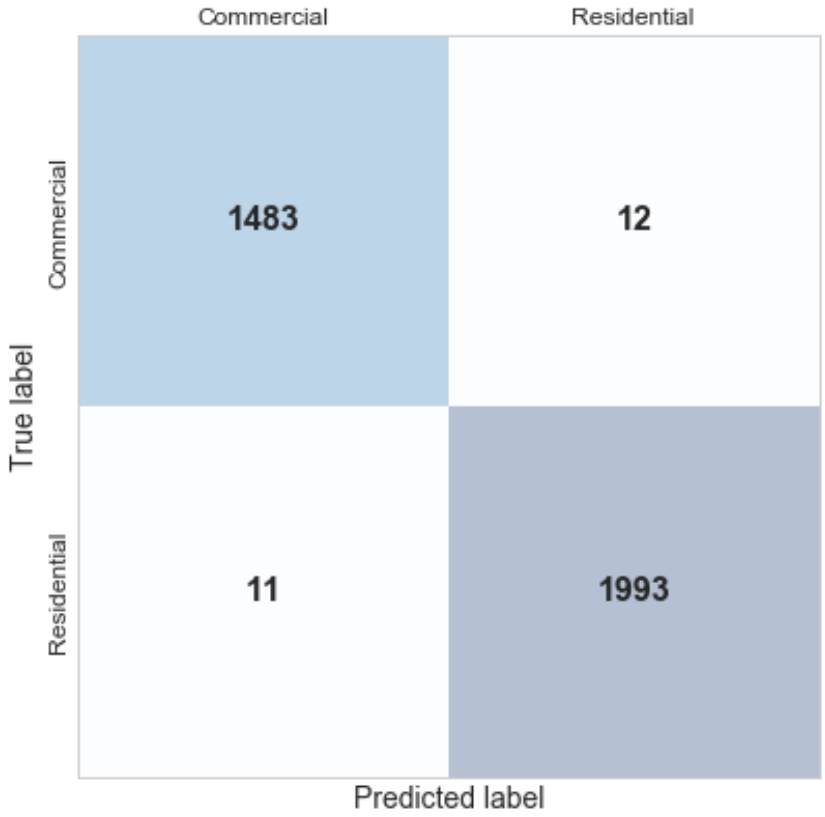




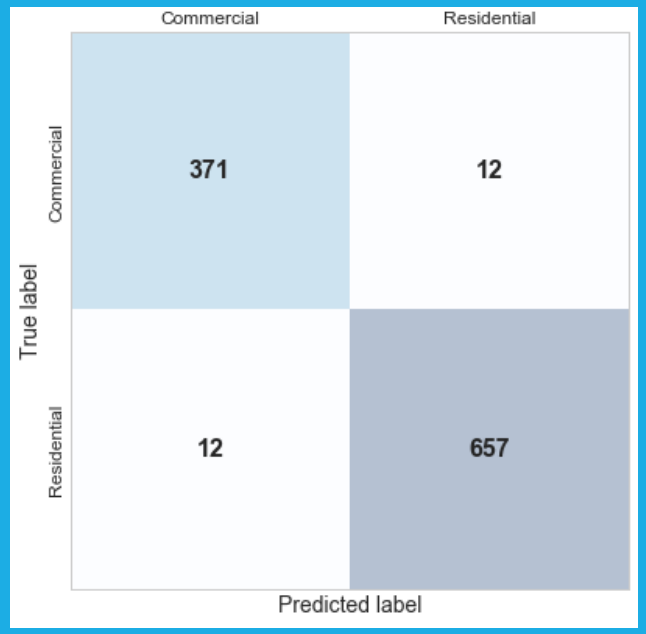
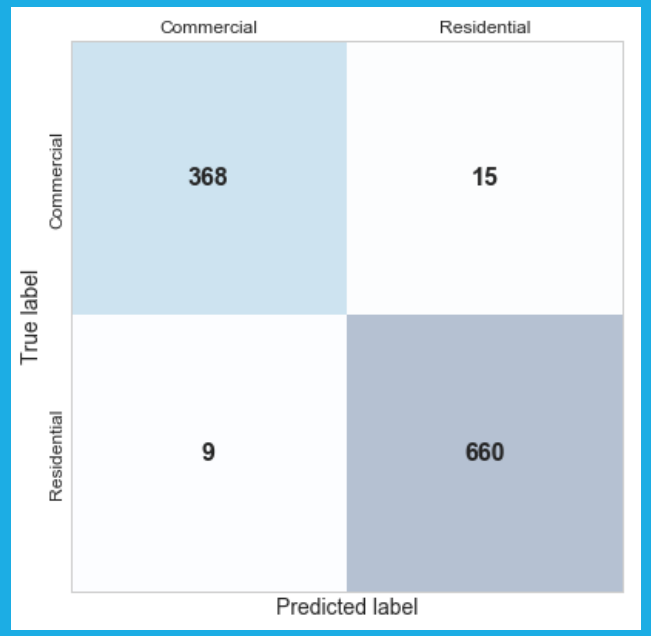
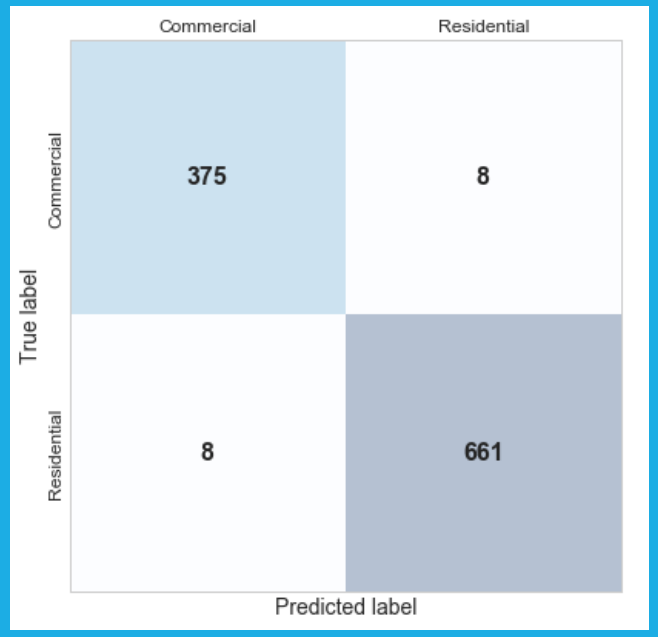
Model Analysis

- Multilayer Perceptron outperforms Random Forest





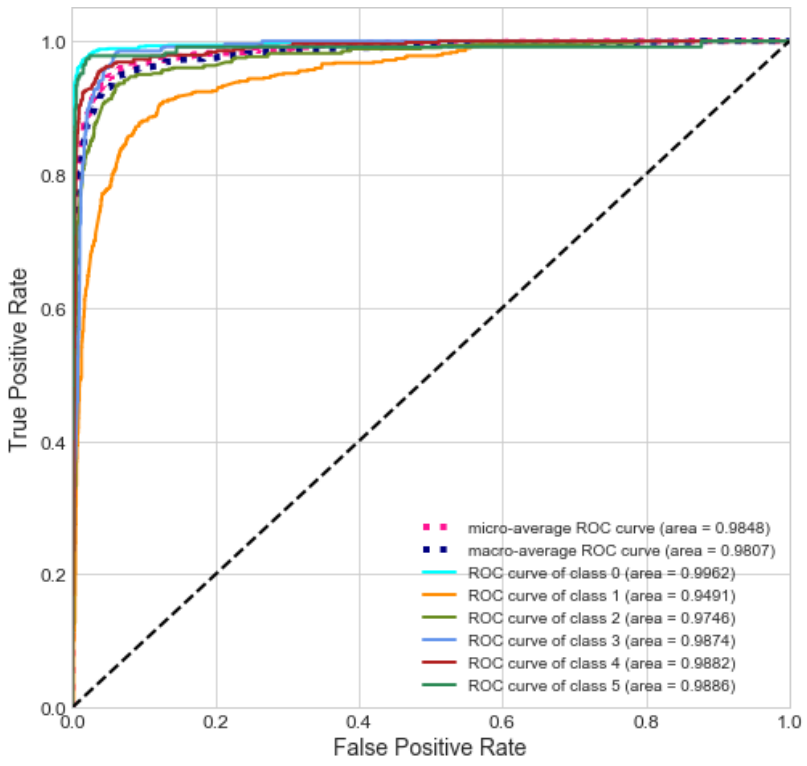
Model Summary

	Logistic Regression	Random Forest	Multilayer Perceptron
Test Accuracy	99.34%	97.97%	99.34%
Validation Accuracy	97.71%	97.71%	98.47%
Confusion Matrix of model on validation set			

Next Steps

Logistic Regression Multiclass Classification | Initial Test Accuracy: 86.04%

True label \ Predicted label	New	Existing	Alteration	Repair	Other	Addition
New	955	10	2	2	3	1
Existing	76	589	35	38	21	3
Alteration	38	54	488	14	2	2
Repair	20	29	7	408	2	0
Other	20	28	0	1	424	2
Addition	18	2	5	0	0	200



Next Steps

- Continue to fine tune and evaluate models and model performance
- Simplify code and isolate each model into its own notebook
- Save models for quicker analysis
- Further develop models for Job Type multiclass classification
- Create a project proposal and deck for presentation to city stakeholders
- Setup and meet with city stakeholders

Appendix

Data Source:
https://data.montgomeryal.gov/Permits/Building-Permit-2014-Present-Download-/qvzc-ejq2

Project Repository:
https://git.generalassemb.ly/bdub595217/project-final

Other Sources:
https://scikit-learn.org

https://machinelearningmastery.com

Zheng, Alice, and Amanda Casari. "Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists." O'Reilly, 2018.

Géron, Aurélien. "Hands-On Machine Learning with Scikit-Learn and TensorFlow." O'Reilly, 2017.