Regression and Classification Results

Machine Learning Project

Szymon Socha

June 6, 2022

Regression

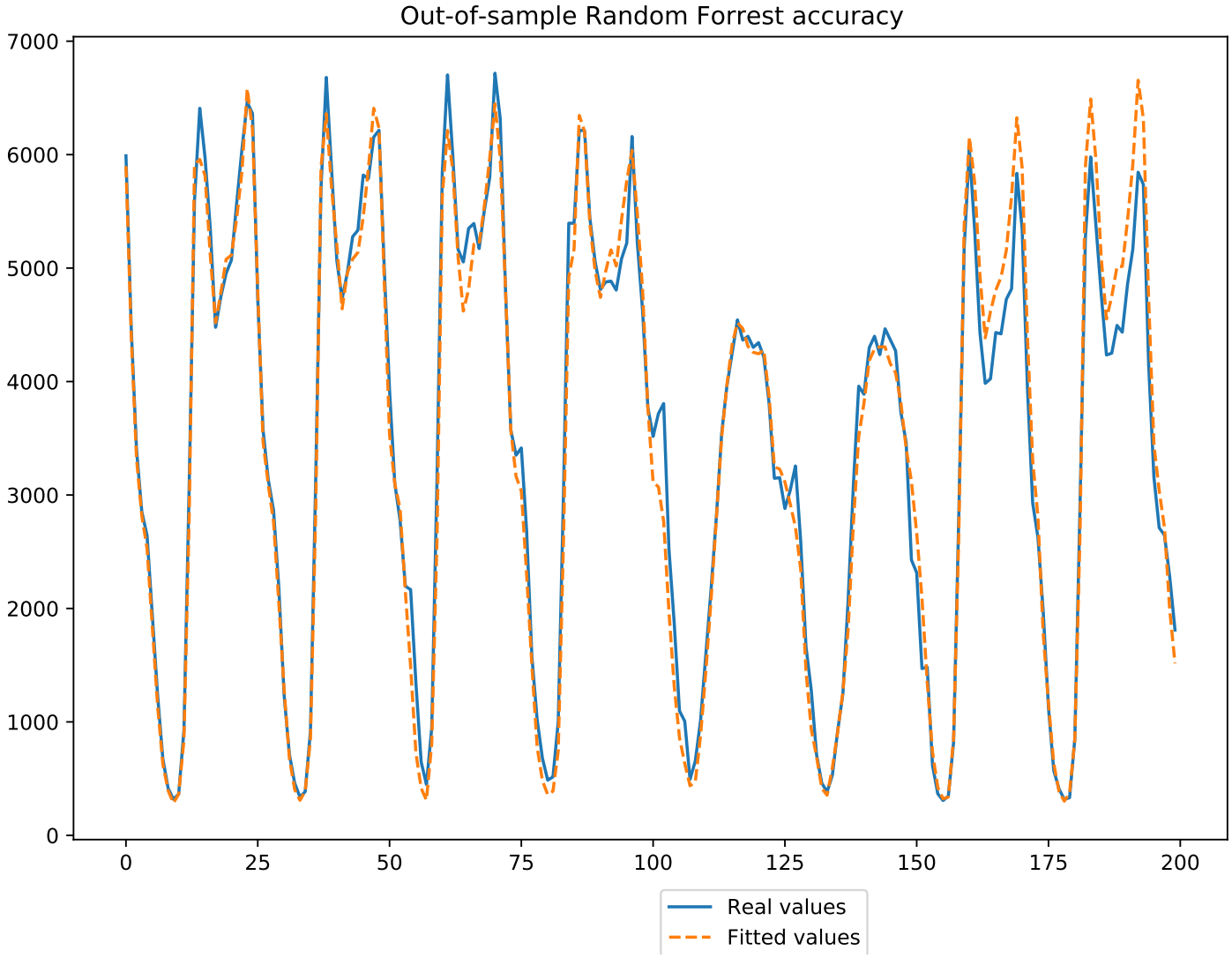
Algorithms considered:

- Linear Regression, Lasso, Ridge, Elastic Net
- **Random Forrest Regressor**
- k-Nearest Neighbors Regressor
- Support-Vector Machine

Algorithm selection:

- Manual Cross Validation parameter tuning
- Out-of-sample accuracy comparision of various algorithms
- **Random Forrest Regressor with** `n_estimators=1000`, `max_features=38`, `min_samples_split=19`, `max_depth=35`, `min_samples_leaf = 3`

Regression



Expected value of **MAPE: 178%**

Classification

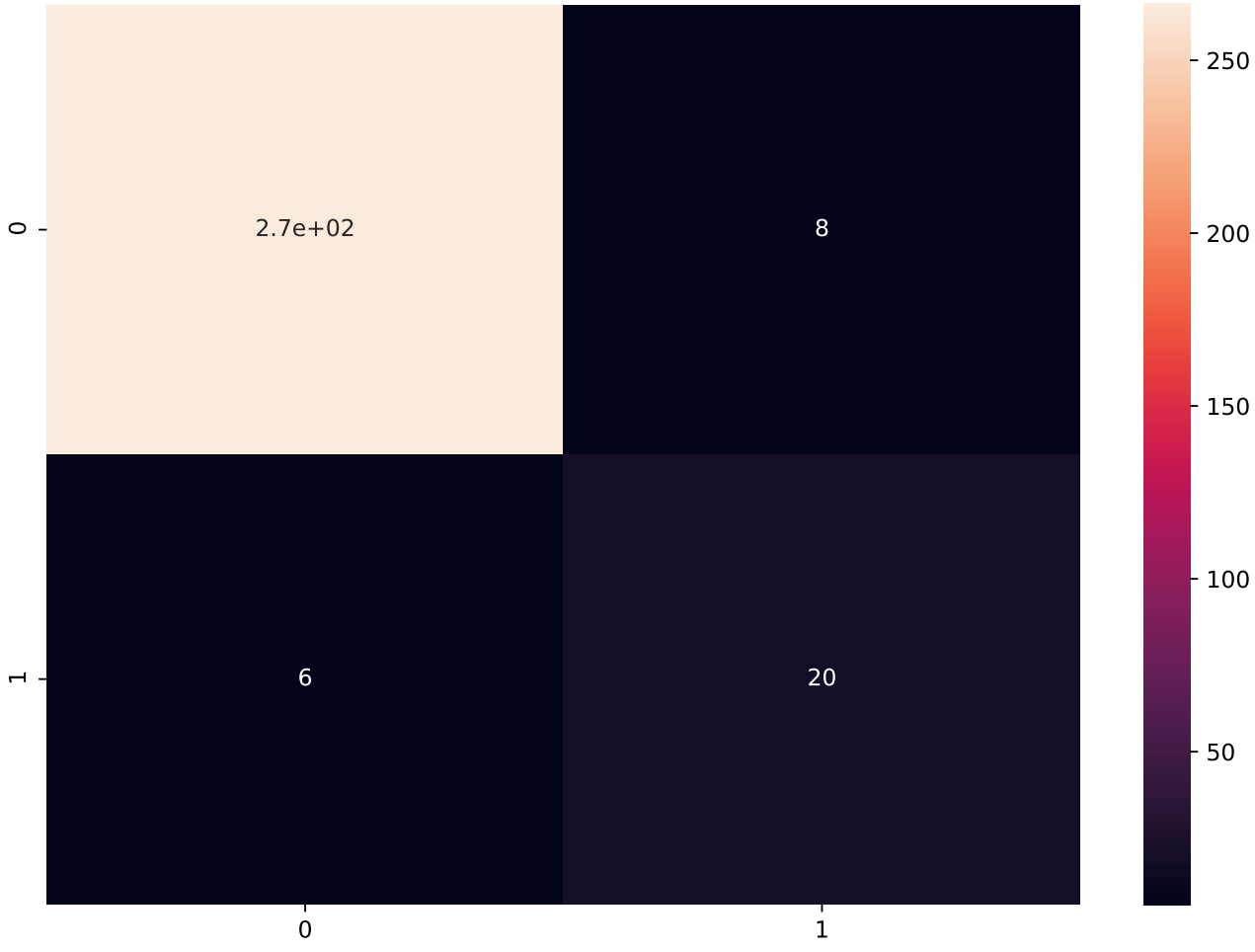
Algorithms considered:

- Random Forrest
- **k-Nearest Neighbors Classifier**
- Logistic Regression
- Support Vector Classifier

Algorithm selection:

- Balancing data with SMOTE+Tomek (reduce recall, increase precision)
- Manual Cross Validation parameter tuning
- Out-of-sample accuracy comparision of various algorithms
- **k-Nearest Neighbors Classifier with `n_neighbors=2`**

Classification



Expected value of `balanced accuracy`: **88%**