Online Shopping Intention and Customer Classification

An analysis by Kai Tamashiro

Can companies use machine learning to increase revenue from both online sales and passively through ads?

Finding the best model for revenue classification

```
sm = SMOTE(random_state=42)
X_trainOS, y_trainOS = sm.fit_resample(X_train, y_train)
Counter(y_trainOS)
```

```
Counter({False: 8307, True: 8307})
```

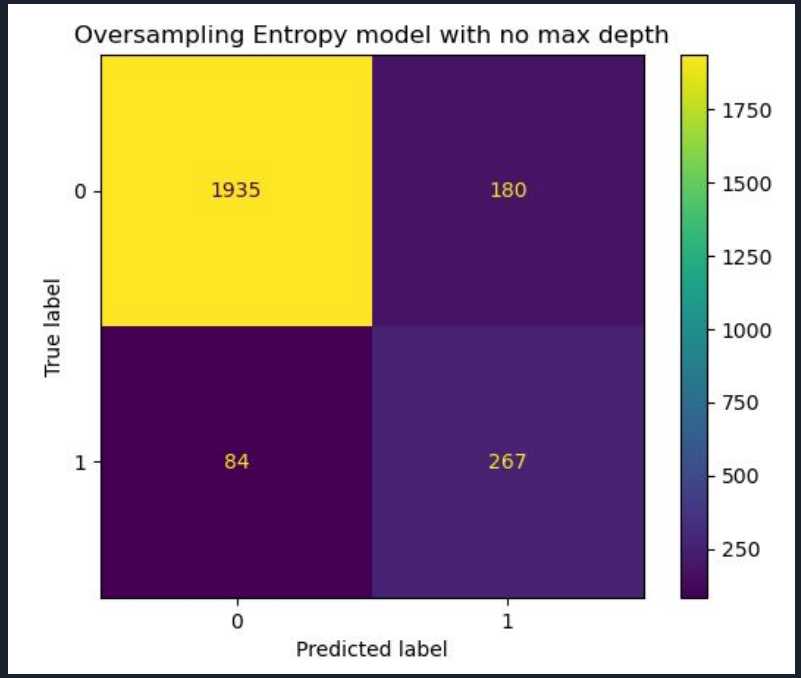
Over sampling to create equal classes

Random Forest: Entropy

Avg precision: .91

Avg recall: .89

F1-score: .90

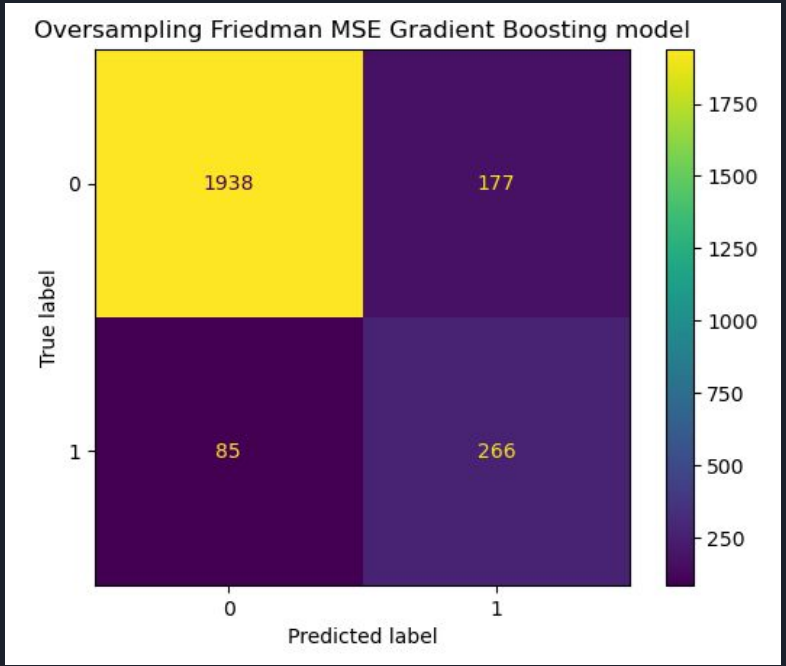


Gradient Boosting

Avg precision: .91

Avg recall: .89

F1-score: .90



Finding the best model, part 2

```
nm = NearMiss()
X_trainUS, y_trainUS = nm.fit_resample(X_train, y_train)
Counter(y_trainUS)
```

```
Counter({False: 1557, True: 1557})
```

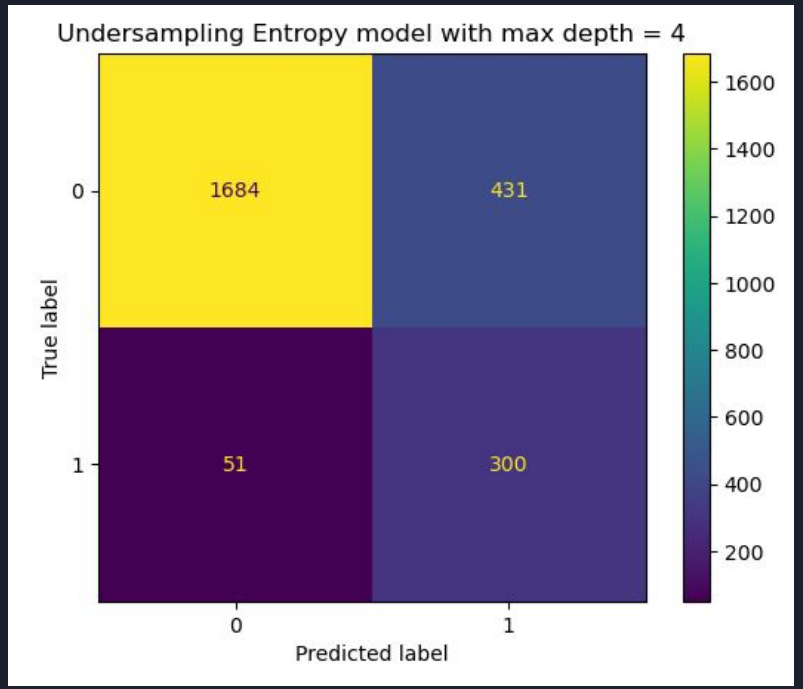
Under sampling to create equal classes

Random Forest: Entropy with max depth

Avg precision: .89

Avg recall: .80

F1-score: .83

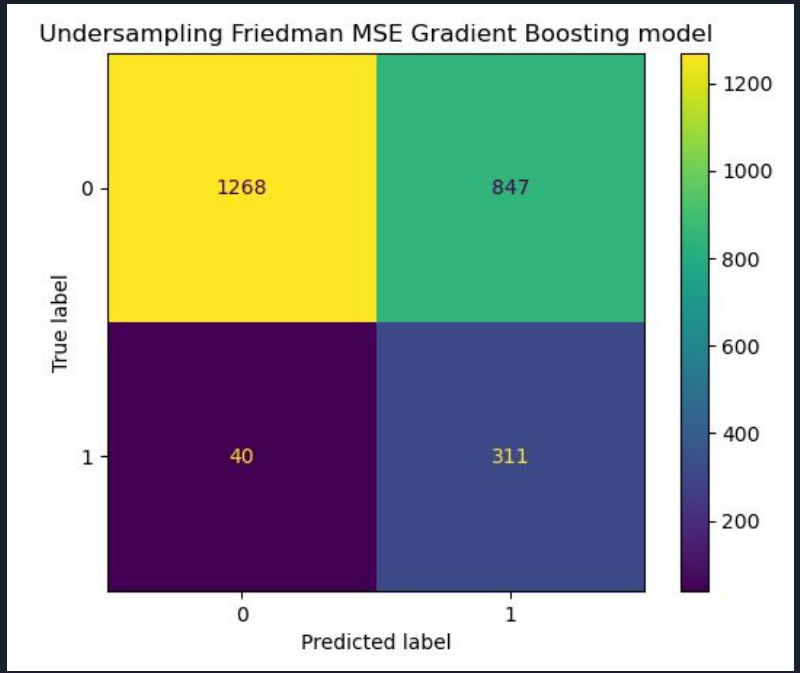


Gradient Boosting, part 2

Avg precision: .87

Avg recall: .64

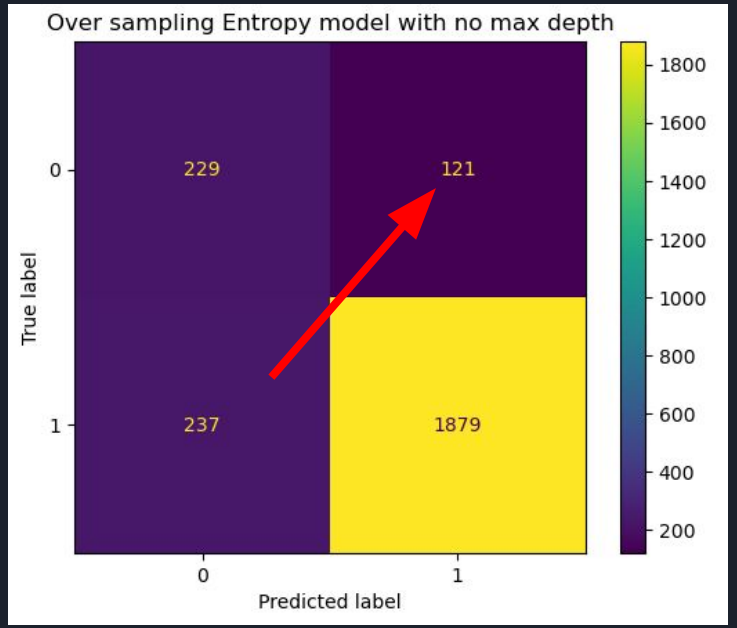
F1-score: .69



Finding the best model for customer classification

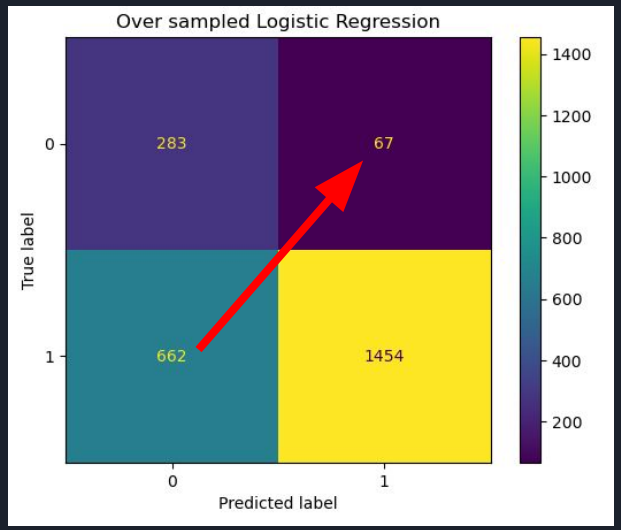
Same steps as revenue classification

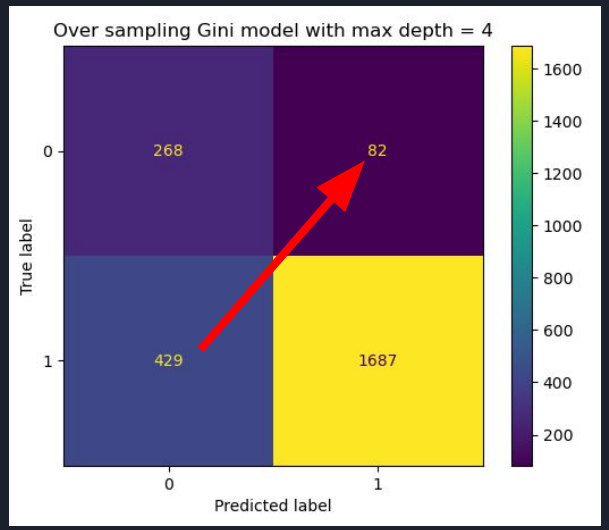
Smallest number of false positives

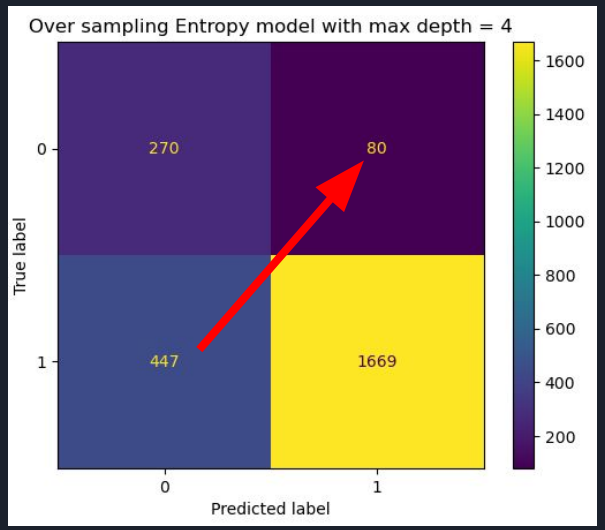


Logistic Regression?

Lowest false positives

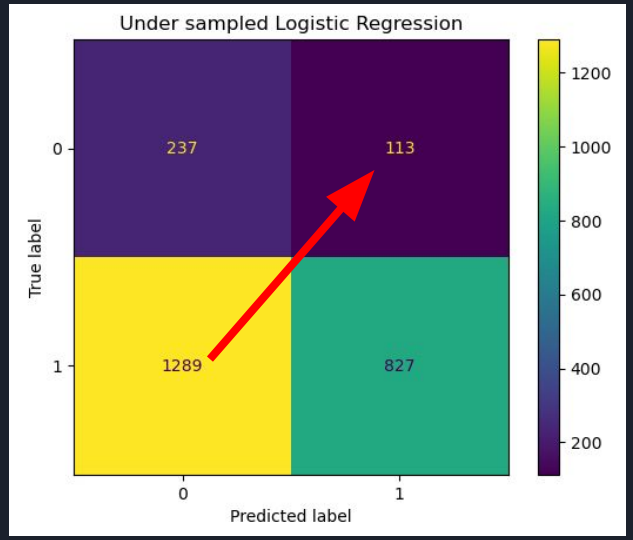


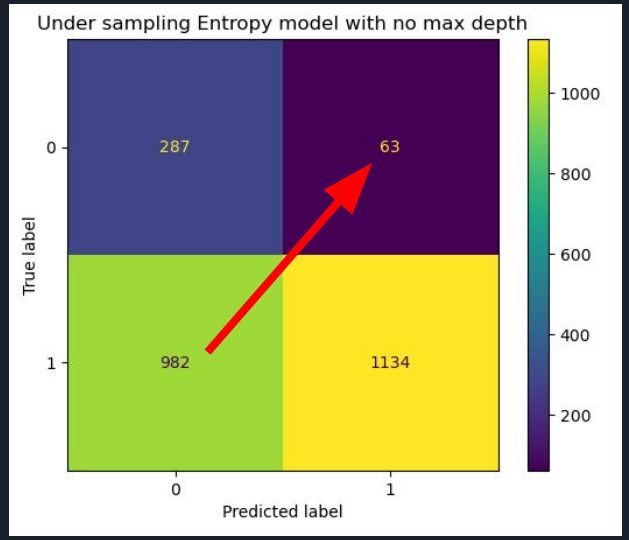


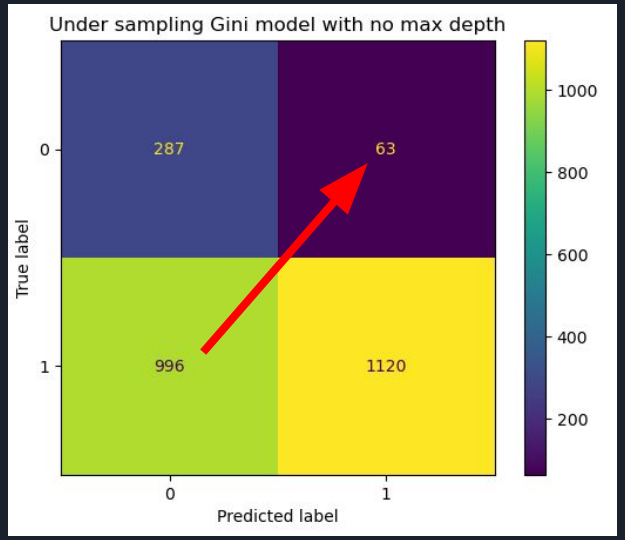


Random Forests Triumph

Under sampling to test flexibility







Next steps

Implement Entropy Random Forest Model

Tweak model to appropriate max depth

Push pop-ups to customers as they are browsing the site

Invest in premium ad network

Passive revenue from ad impressions