<> **Code**   ⓘ Issues   ⑊ Pull requests   ▶ Actions   ▦ Projects   📖 Wiki   🛡 Security   📈 Insights   ⚙ Settings

ⴷ main ▾   Go to file   Add file ▾   ⬇ Code ▾

About ⚙

👩 **carlytsu** Update README.md ...		5 days ago	🕑 **5**
📁 .ipynb_checkpoints	removed lab notebooks		5 days ago
📁 data	for initial project review, sub mvp		7 days ago
📁 img	finalized notebook and presentation, all files		5 days ago
📄 EcommerceClassificatio...	finalized notebook and presentation, all files		5 days ago
📄 Final_Project3_Noteboo...	finalized notebook and presentation, all files		5 days ago
📄 Project_3_actual.ipynb	finalized notebook and presentation, all files		5 days ago
📄 README.md	Update README.md		5 days ago

No description, website, or topics provided.

📖 Readme

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%

☰ **README.md** ✏

by Carly Tsuda, for the Flatiron School Data Science Program, Phase 3

Business problem:

The project in this repository creates classifier models to attempt to determine which sessions on an ecommerce website resulted in sales. The imagined audience is a web development startup (such as SquareSpace) which hopes to create new features to help clients get the most out of their online stores.

Data:

The dataset comes from the UCI machine learning repository. It contains data on 12,330 unique sessions on an unspecified ecommerce site from 2018.

Methods:

I used three classification models (Logistic Regression, Decision Tree, and Random Forest) to determine predict which sessions ended in sales. For the Decision Tree and Random Forest, I also implemented a Grid Search to find the optimal combination of hyperparameters. I implemented the entire process in a data science Pipeline to streamline preprocessing.

Results:

Each of the three classification models quickly converged at similar metrics. Both the Decision Tree and Random Forest were overwhelmingly weighted toward a single feature: average Page Value (determined by Google Analytics).

Classifier Prediction Success Rates
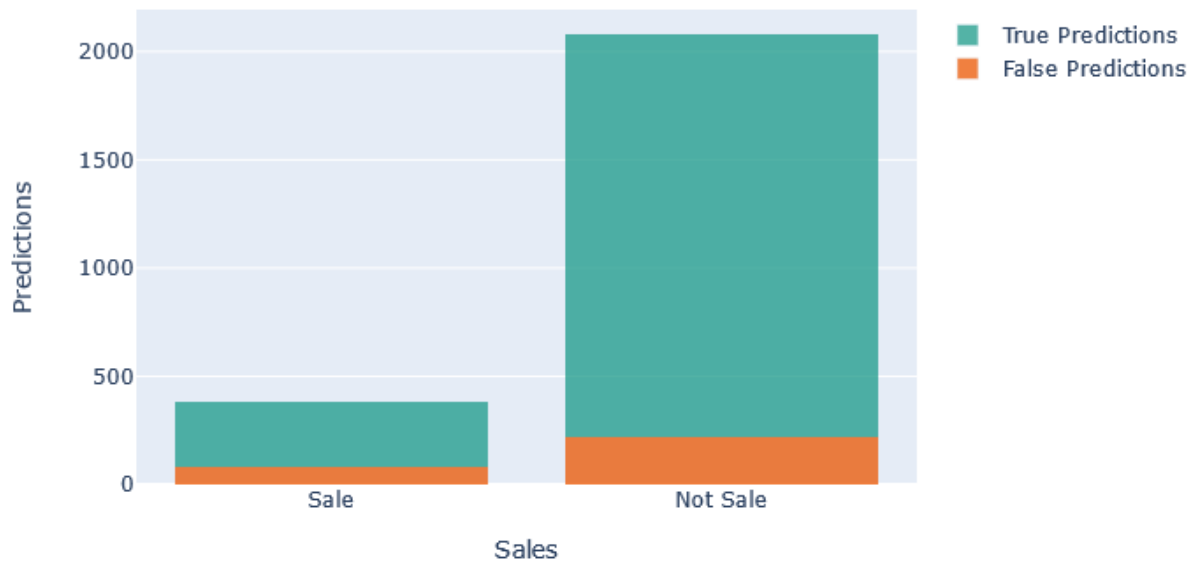
Model	Accuracy (Train)	Accuracy (Test)	Precision (Train)	Precision (Test)	Recall (Train)	Recall (Test)
Logistic Regression	0.85	0.84	0.52	0.50	0.77	0.77
Decision Tree	0.90	0.89	0.91	0.61	0.88	0.76
Random Forest	0.90	<0.88	0.89	0.58	0.91	0.79

Reccomendations:

The strength of the Page Values section affirms the efficacy of analytic data. Recommendations included a focus on empowering site owners to leverage analytic data to enhave existing business understanding.

For further information:

Please review the narrative of the analysis in my jupyter notebook and review my non-technical presentation.

Repository structure:

```
├── README.md
├── Project_3_actual.ipynb                    <- Jupyter Notebook
presentation
├── EcommerceClassification_Nontechnical.pdf  <- Non-technical
presentation
├── Final_Project3_Notebook.pdf               <- PDF of Jupyter
Notebook
```

```
        └── img                              <- directory
containing visualizations and images for presentation
            └── images
        └── data                             <- contains original
dataset and data dictionary
```