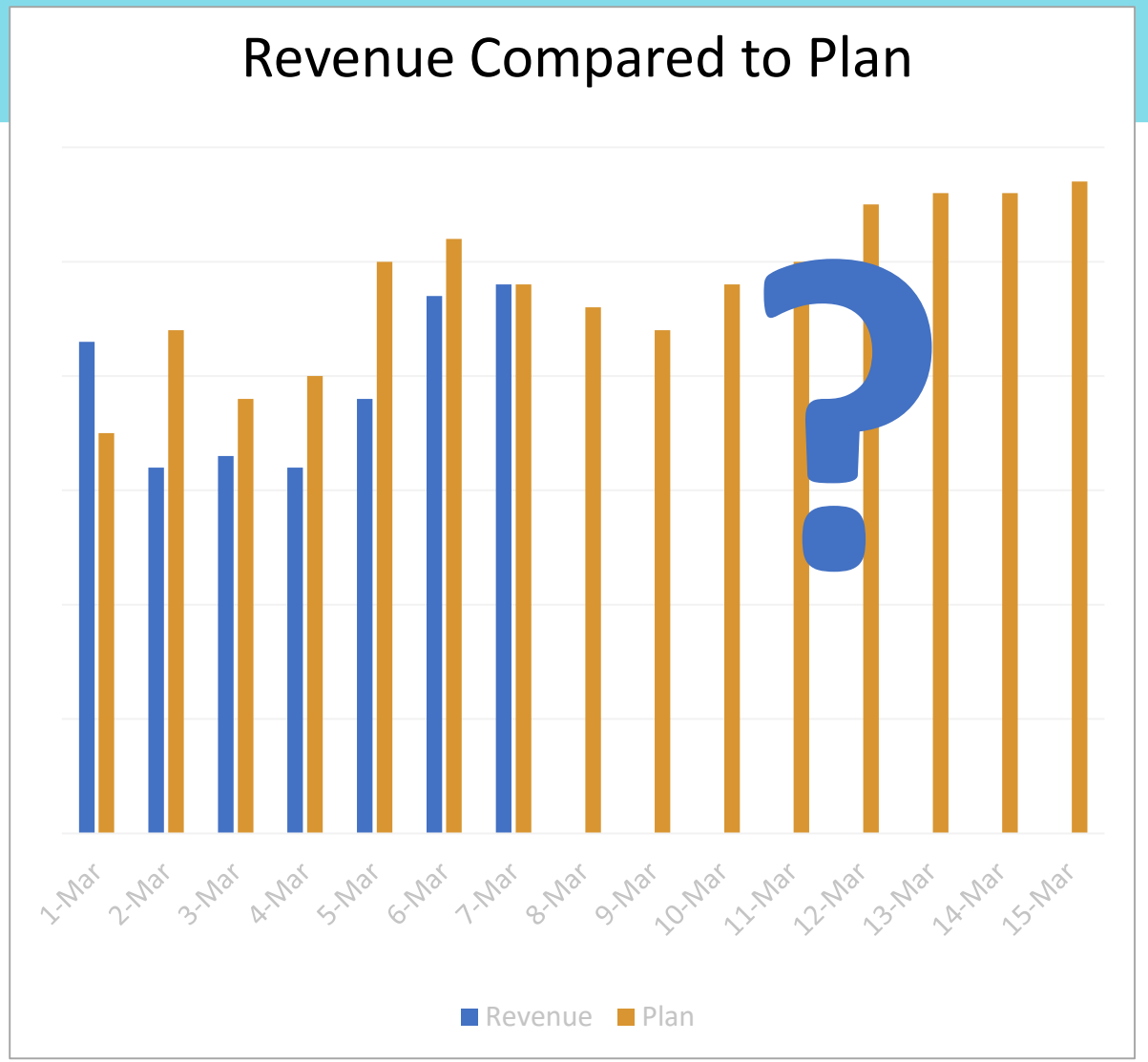
BlueMoon/ Digital™

Predicting if Daily Revenue will Reach Daily Forecasted Revenue Using Machine Learning

by: Julianna Renaud

The Project

To build a prediction model for eCommerce retailers that helps them understand if they are likely to reach their daily revenue plan numbers.



BlueMoon/ Digital

The Client

A women's fashion retailer selling a variety of product categories such as clothing, shoes, jewelry and handbags.

**For confidentiality purposes, the name of the client will remain anonymous.*

BlueMoon/ Digital

The Data

- Historical eCommerce Website Performance Data (Google Analytics)
- Historical Paid Advertising Spend Data (provided from the client)
- The Marketing Calendar (provided from the client)
- Daily Plan Data (provided by the client)

Google Analytics

BlueMoon/ Digital

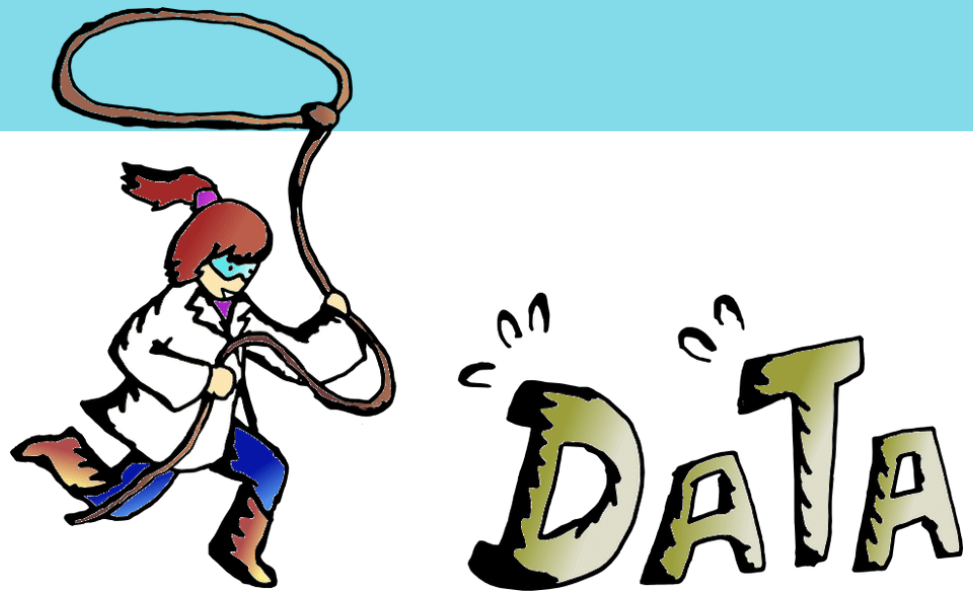
The Methodology

The project was treated as a supervised learning classification problem. The model was built to predict the binary variable ‘Goal Reached’.

Date	Goal Reached
March 1	Y
March 2	N
March 3	N
March 4	N
March 5	N
March 6	N
March 7	Y
March 8	Y
March 9	Y
March 10	Y
March 11	N
March 12	N
March 13	Y
March 14	N
March 15	Y

BlueMoon/ Digital

Data Wrangling



- Adjust Formatting (date & seconds)
- Handle Missing Data
- Shift Historical Data Forward
- Create Binary Fields Based on Marketing Event Data & Historical Achievement/Non Achievement of Plan
- Scale the Data

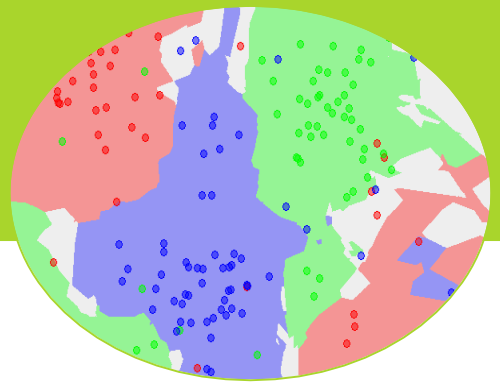
BlueMoon/ Digital

Exploratory Data Analysis

- Identify the benchmark for prediction (the percentage of days that the client currently achieves goal without)
- Analyzed data to identify trends
- Identified and subsequently removed outliers
- Explored the impact of marketing events on revenue

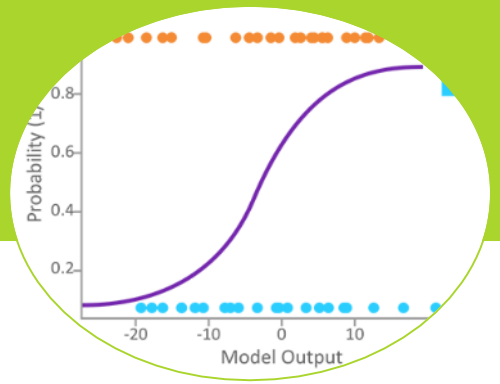
BlueMoon/ Digital

The Models



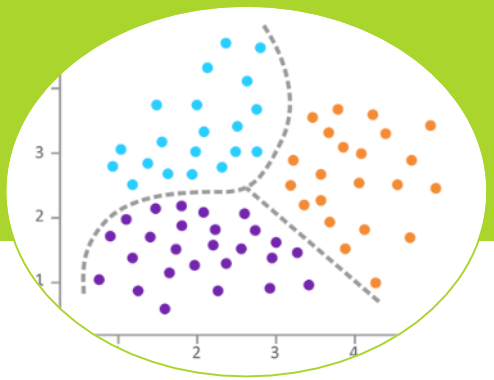
K Nearest Neighbors

Non-parametric method used for classification. Will classify based on a majority vote of it's neighbors.



Logistic Regression

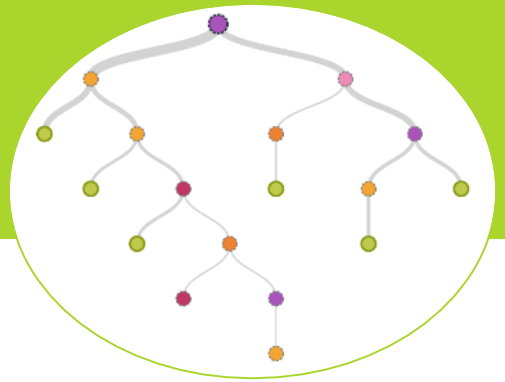
The "go-to" method for binary classification problems.



Bernoulli Naïve Bayes

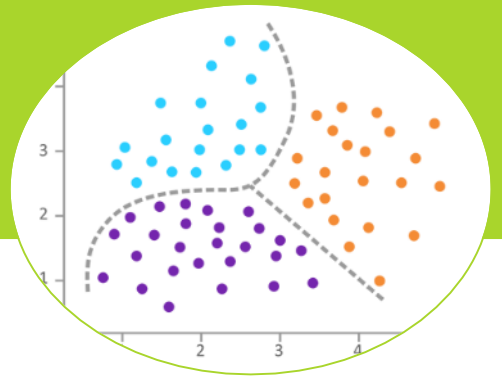
Classifier that is designed for working with independent binary/ Boolean features.

The Models



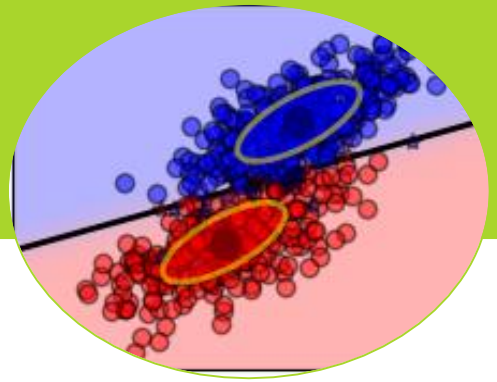
Decision Tree

A “flowchart-like” structure. Each node represents a “test” on a feature and each branch represents the outcome of the test.



Gaussian Naïve Bayes

Assumes a normal distribution. Typically used for working with continuous data.



Linear Discriminant Analysis

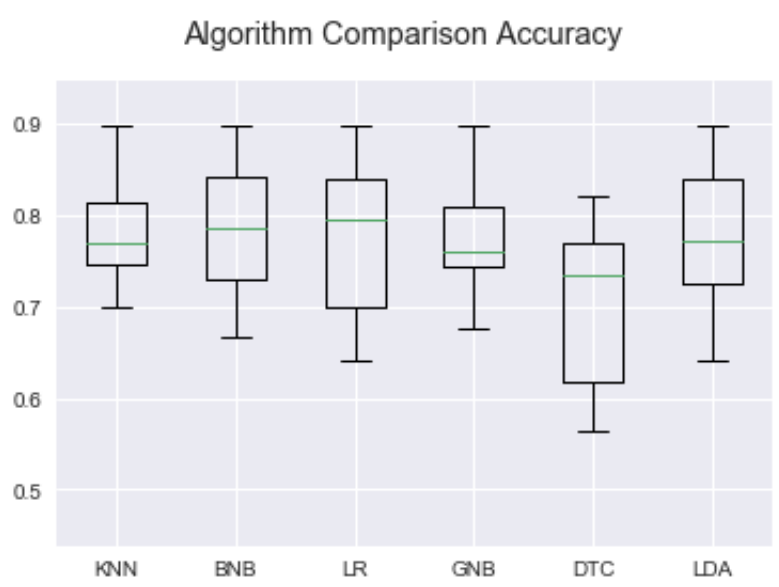
Method used for finding a linear combination of features that characterize two or more classes.

Model Evaluation

Model	Precision Average	Recall Average	F1-Score Average	Accuracy Score	ROC AUC
K Nearest Neighbor	0.79	0.80	0.79	0.75	0.74
Logistic Regression	0.78	0.78	0.78	0.77	0.72
Bernoulli NB	0.81	0.81	0.81	0.78	0.76
Decision Tree	0.61	0.57	0.58	0.70	0.68
Gaussian NB	0.80	0.80	0.80	0.78	0.77
Linear Discriminant Analysis	0.81	0.81	0.80	0.78	0.79

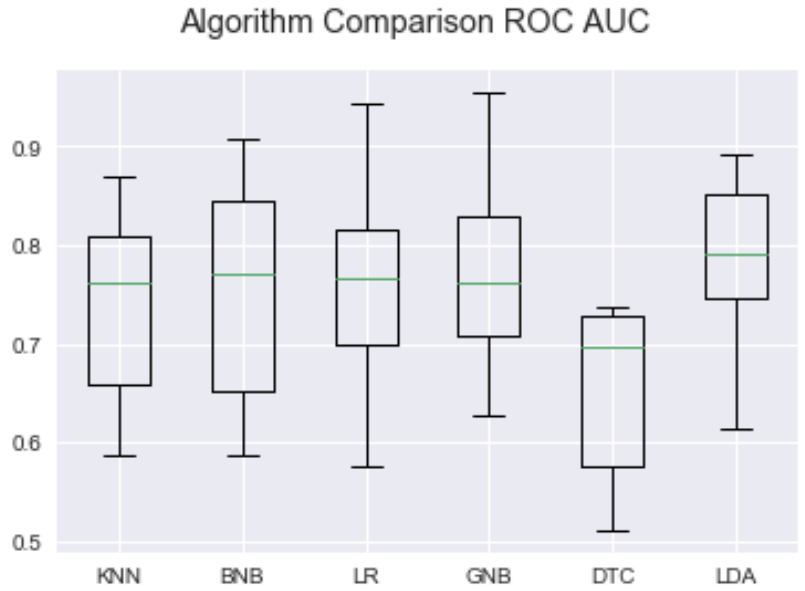
Model Evaluation

Accuracy




```
KNN: 0.753141 (0.109882)
BNB: 0.783846 (0.069194)
LR: 0.773718 (0.081124)
GNB: 0.778910 (0.076573)
DTC: 0.702244 (0.086164)
LDA: 0.778654 (0.077992)
```

ROC AUC




```
KNN: 0.741652 (0.091609)
BNB: 0.751358 (0.110878)
LR: 0.760065 (0.109065)
GNB: 0.777721 (0.101402)
DTC: 0.653324 (0.087410)
LDA: 0.789050 (0.078256)
```

BlueMoon/ Digital

Next Steps

- Acquire additional historical plan and paid media spend data
- Possibly add additional data that documents any site optimization efforts that may have impacted (or could impact) eCommerce performance
- Initiate automated processes for data ingestion into the model
- Leverage the client's BI platform to visualize the model's predictions

Client Recommendations

- Recommendation 1: That we continue to move forward with the project by automating data ingestion into the model.
- Recommendation 2: That we then run multiple models (Bernoulli NB, Linear Discriminant Analysis and Gaussian NB) simultaneously on live data to gauge which model will make the best predictions on future data.
- Recommendation 3: We begin to utilize model predictions to make marketing initiative updates to improve revenue performance on days not predicted to reach plan.

THANK YOU

Blue Moon Digital, Denver
1512 Larimer Street, Suite 800
Denver, Colorado 80202

Julianna Renaud

Web Analytics Strategist

Charlotte Werger, PhD

Springboard Mentor

BlueMoon/ Digital | Springboard