STAT 451 FINAL PROJECT

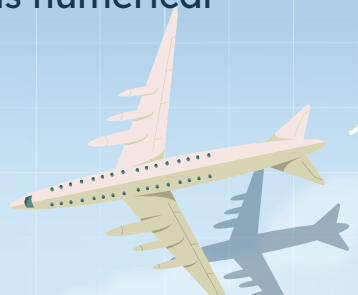
Group 18 - Kellan Meyer, Lili McGuffey,
Evelyn Tollar, Gabe Walter, Isabelle Gallucci

OUR DATA

We analyzed Invistico airline satisfaction data

	satisfaction	Customer Type	Age	Type of Travel	Class	Flight Distance	Seat comfort	Departure/Arrival time convenient	Food and drink	Gate location	...	Online support	Ease of Online booking	On-board service	Leg room service	Baggage handling	Checkin service	Cleanliness	Online boarding	Departure Delay in Minutes	Arrival Delay in Minutes
43191	dissatisfied	disloyal Customer	36	Business travel	Eco	2171	1	1	1	4	...	2	1	2	1	4	2	3	1	0	8.0
125240	satisfied	Loyal Customer	30	Business travel	Business	2632	4	4	4	4	...	5	5	4	2	5	3	5	5	0	0.0
28491	dissatisfied	Loyal Customer	25	Personal Travel	Business	2487	3	5	2	3	...	5	5	4	5	4	5	5	5	0	3.0
124201	satisfied	Loyal Customer	29	Business travel	Eco	2784	5	3	3	3	...	5	5	3	2	4	5	3	5	0	18.0
39421	satisfied	Loyal Customer	29	Personal Travel	Eco	1459	5	4	5	5	...	4	4	4	5	4	4	4	4	0	0.0

** Made satisfaction, customer type, type of travel, and class columns numerical



OUR QUESTIONS

- **Q1: What factors most affect customer satisfaction with Invistico Airline?**
  - Overall “satisfied” or “dissatisfied”
- Possible Factors:
  - Age
  - Flight class
  - Customer loyalty
  - Type of travel
  - Ratings such as seat comfort, online support, baggage handling, etc.
    - Rating out of 5
  - Departure and arrival delay
- Feature Selection
  - Lasso
  - Select K Best

- **Q2: Based on the most influential factors, which classification model is best suited for predicting customer satisfaction?**
- Possible Models:
  - Logistic regression
  - Knn
  - Svm
  - Decision Tree

SELECTING FEATURES

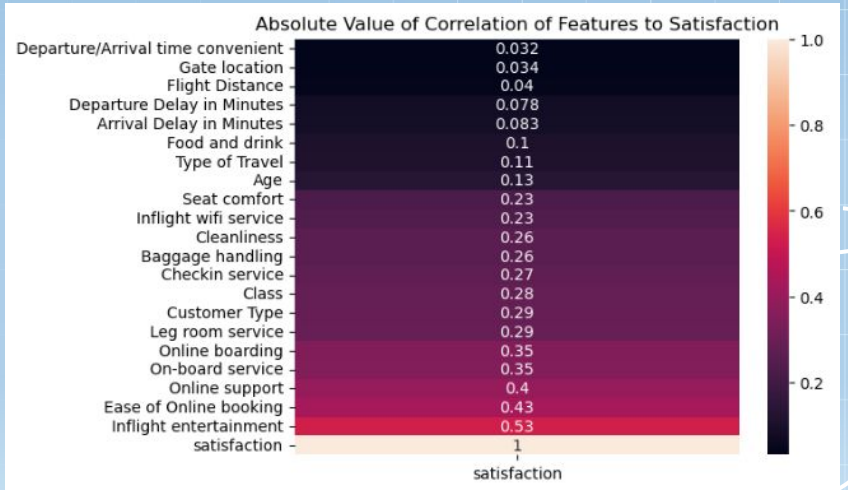
Lasso also works well with multicollinearity

```
Class -0.27934
Flight Distance -0.00031
Departure/Arrival time convenient -0.0867
Gate location -0.08
Inflight entertainment 0.36198
Ease of Online booking 0.13501
Arrival Delay in Minutes -0.00518
```

Logistic Lasso (C = 0.003)

Correlation to Satisfaction

```
['Customer Type',
 'Inflight entertainment',
 'Online support',
 'Ease of Online booking',
 'On-board service',
 'Leg room service',
 'Online boarding']
```



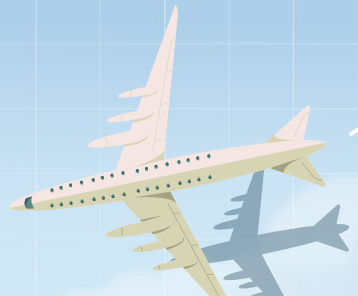
SelectKBest

HIGHLIGHT OF FEATURE SELECTION

Based on our Lasso feature selection we found that the top seven:

1. Class
2. Flight Distance
3. Departure/Arrival Time
4. Gate Location
5. Inflight Entertainment
6. Ease of Online Booking
7. Arrival Delay

Were the top 7 most important factors when considering satisfaction



SUMMARY OF METHOD

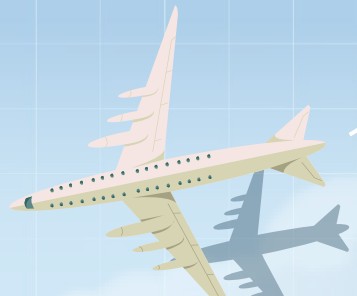
Grid Search CV for Select K Best

- Pick Model + Parameters:
  - SVM (Kernel, C)
  - Logistic Regression (C)
  - Decision Tree Classifier (Criterion, Max Depth)
  - KNN (# neighbors)

Manual Search for Lasso

- Used best score from these four models and variations

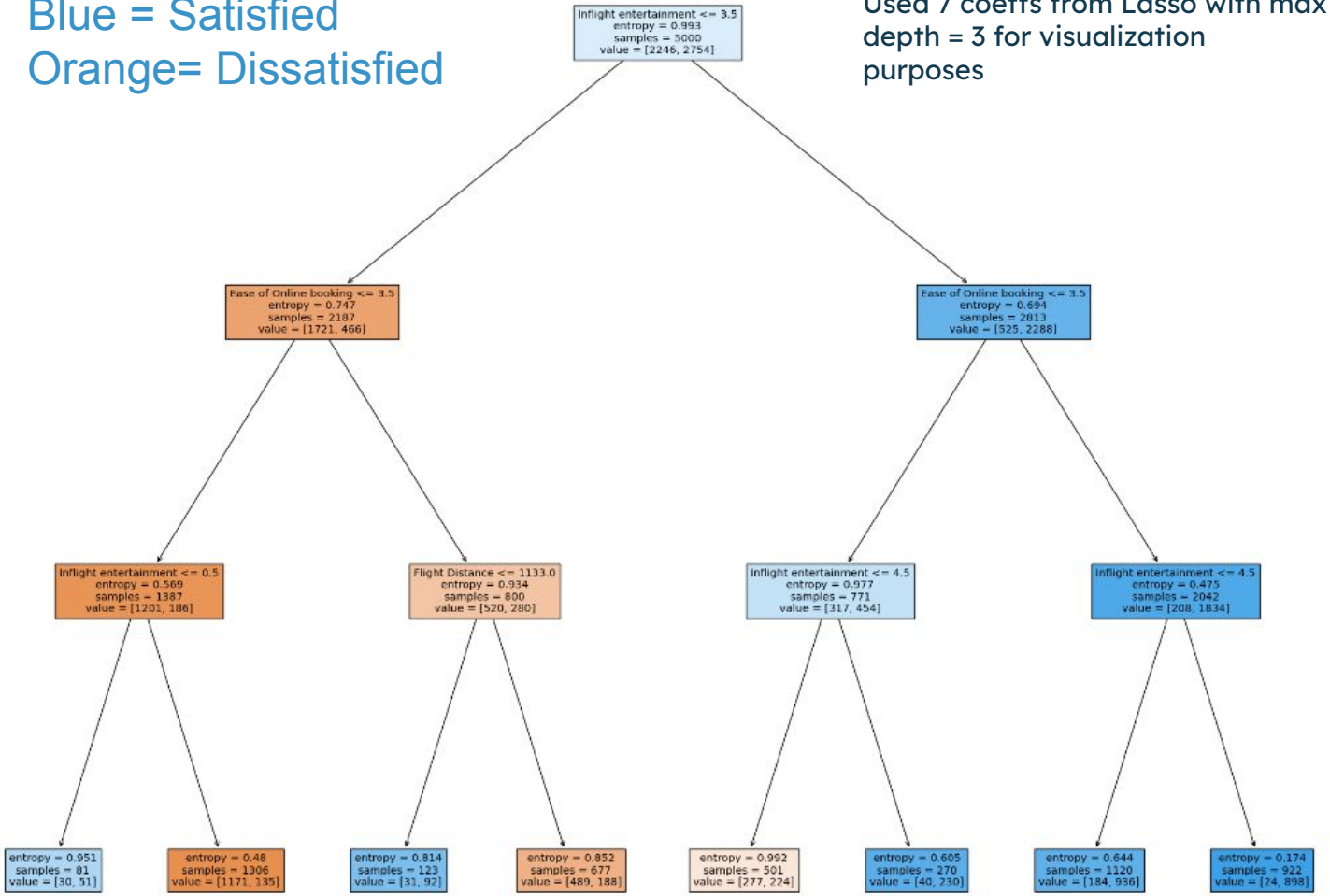
Chosen Features = Lasso, because higher score on unseen test data



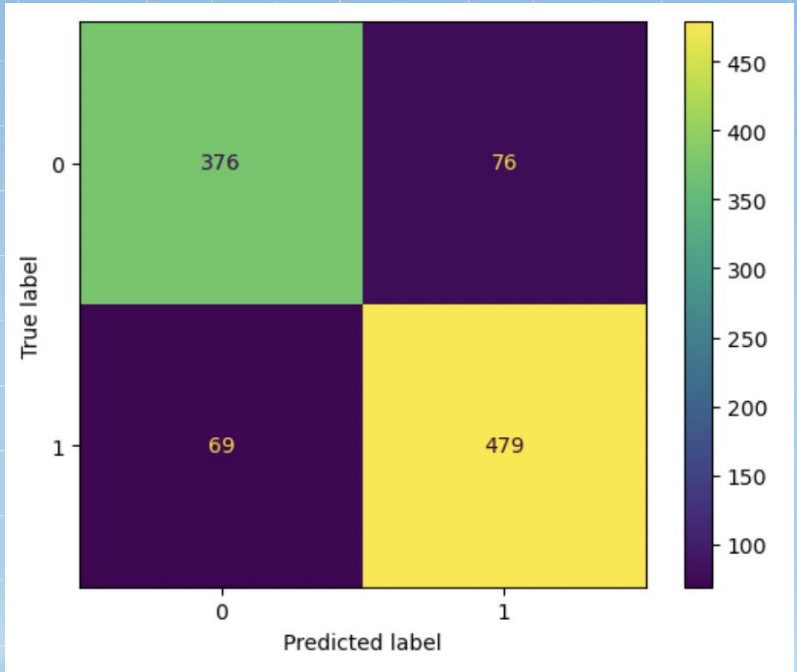
MODEL

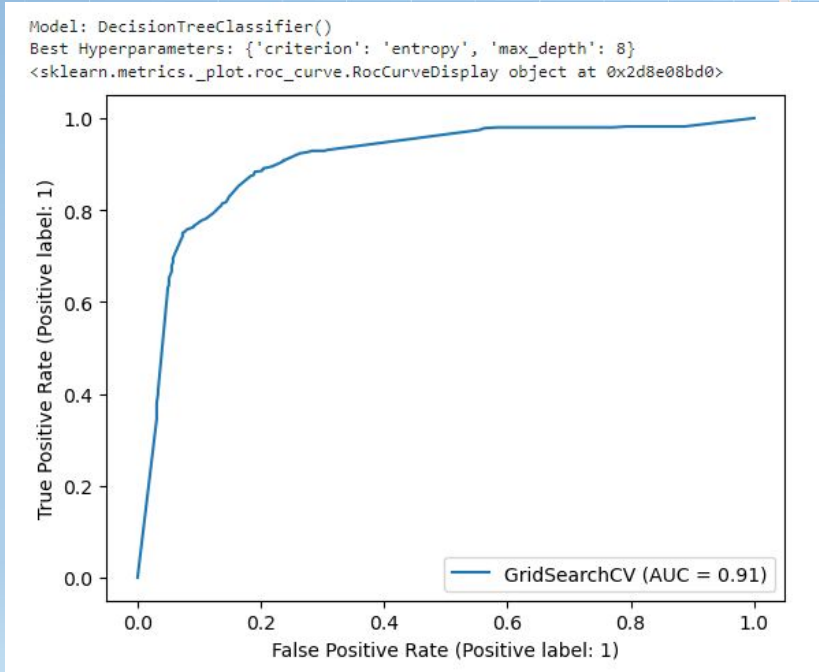
Blue = Satisfied
Orange= Dissatisfied

Used 7 coeffs from Lasso with max depth = 3 for visualization purposes



EVALUATION OF MODEL





Accuracy on test data = 0.855

Accuracy on validation data = 0.870

HIGHLIGHT OUTCOMES

We found that the Decision Tree model was the best model for predicting satisfaction.

With the Decision Tree model using the best 7 coefficients from the Logistic Lasso we got a score of 0.87 on validation data 0.855 on the test data with max depth = 8.

If an airline wanted to increase sales they could apply this model to predict the percentage of satisfied passengers for the coming year, and then set goals to beat