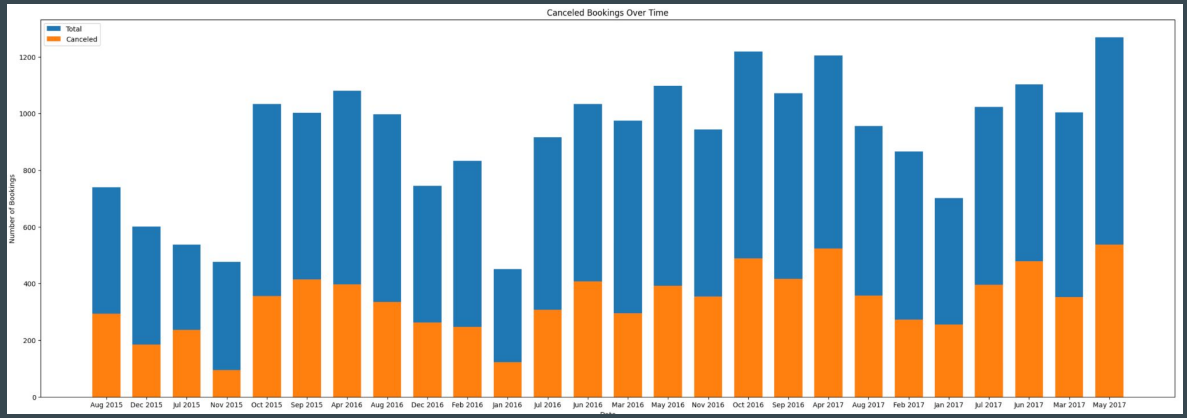
Predicting Hotel Reservation Cancellations

Nimra Shakoor (ns924), Preeti Uppuluri (pnu3), Ashley Yu (yy346)

Problem

- Booking cancellations result in losses of profit for hotels
- The percentage of booking cancellations has been increasing over time [1]



- Knowing the likelihood of a hotel booking being canceled might help them decide how to take action in advance, such as through pricing, advertising, or overbooking

Pictured: Canceled bookings per month over time

Techniques

- Models
  - SVM
  - Random Forests
  - Logistic Regression
- Validation
  - Hyperparameter search
  - 3-fold cross validation

Model	Hyperparameter	Values
SVM	C	[0.01, 0.1, 1, 10]
	Kernel	[poly, rbf]
Random Forests	n_estimators	[200, 1000, **2000**]
	max_depth	[10, **50**, 100, None]
Logistic Regression	C	[0.01, 0.1, 1, 10]
	Solver	[lbfgs, liblinear, newton-cg, newton-cholesky, sag, saga]

Results: SVM

- Metrics
  - Accuracy 0.63
  - Precision of 1
  - Recall of 0.63
  - F1 0f 0.78
  - False Negative rate of 1
- All combinations of hyperparameters performed the same; neither the C-value nor the kernel used yielded different results

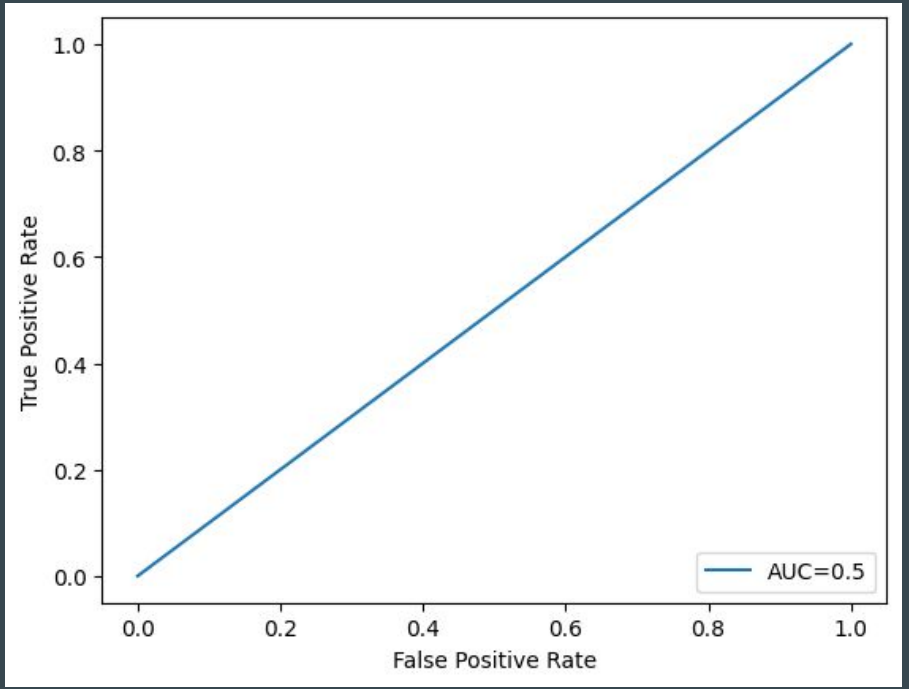
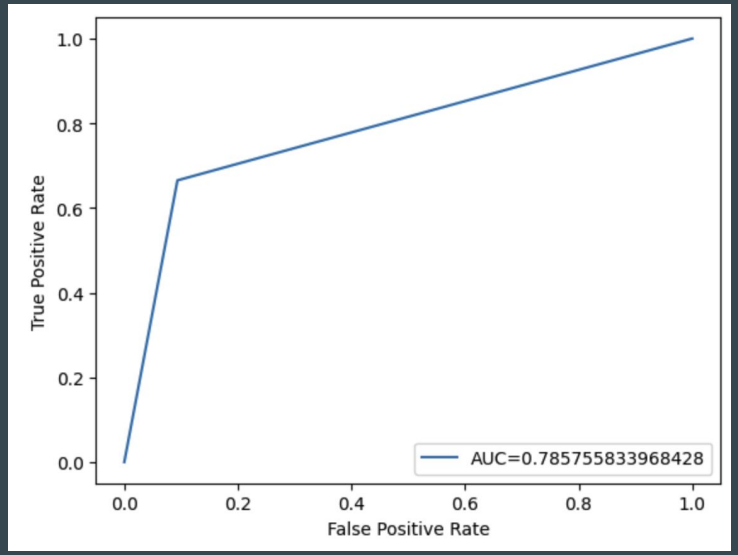


Figure: SVM AUC/ROC Curve

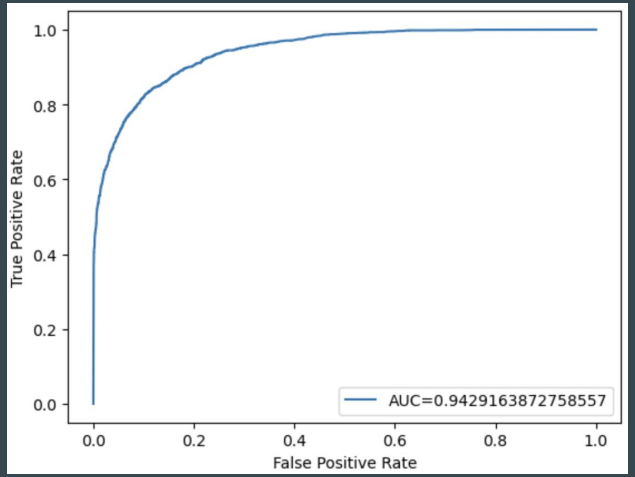
Results: Logistic Regression

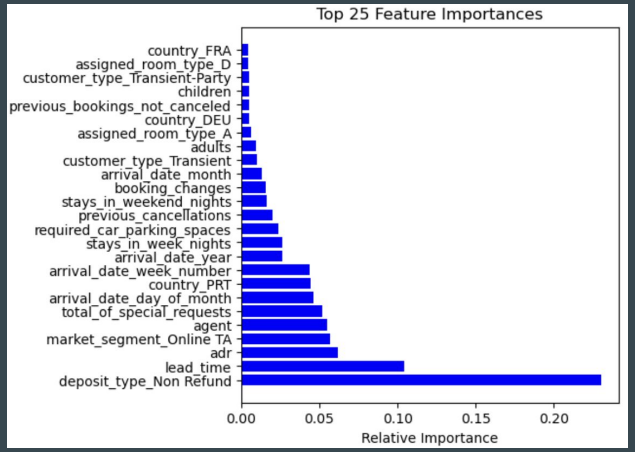
- Average cross-validation score = 0.762
- Optimal hyperparameters
  - C = 10
  - Solver = ‘newton-cholesk’
- Metrics
  - Accuracy = 0.82
  - Precision = 0.91
  - Recall = 0.82
  - F1 = 0.86
  - False Positive Rate = 0.18



Results: Random Forests

- Overall best performing model
- Optimal hyperparameters
  - n_estimators = 2000
  - max_depth = 50
- Metrics
  - Accuracy = 0.87
  - Precision = 0.91
  - Recall = 0.89
  - F1 = 0.90
  - False Negative Rate = 0.11





Conclusion

- Best-performance Model is Random Forest
- Confident with our results
  - Data Quality and Processing
  - Multiple Metrics and Validation Techniques
  - Model Interpretability
- Potential “Weapon of Math Destruction”
  - If being used inappropriately
  - Promote responsible and ethical use of our model

Future Work

- Generalization
  - Compare performance on hotels in vs. outside Portugal
  - How does the model perform on more recent data?

- Different methods
  - Change the framing of the problem: time-series
  - Try more complex models: neural networks

Thank You!

References

[1] P. Delgado, “Cancellations shooting up: Implications, costs and how to reduce them,” May 2016. [Online]. Available: https://www.mirai.com/blog/cancellations-shooting-up-implications-costs-and-how-to-reduce-them/