LUCID SOLUTIONS

Predicting Electric Vehicle Ownership in U.S Market

Riad Rehman

Induranga (Indu) Jayewardene

Lorena Severino

Kwang Heum (Sean) Yeon

Tony Tan

Yong Wu

Agenda

01 **Business Problem**

- ○ Recap of business problem

02 **Model**

- ○ The models , results, and our interpretation

02 **Evaluation**

- ○ The effectiveness of the models & visuals

03 **Conclusion**

- ○ Business implications & deployment of model
- ○ Concerns

Business Problem

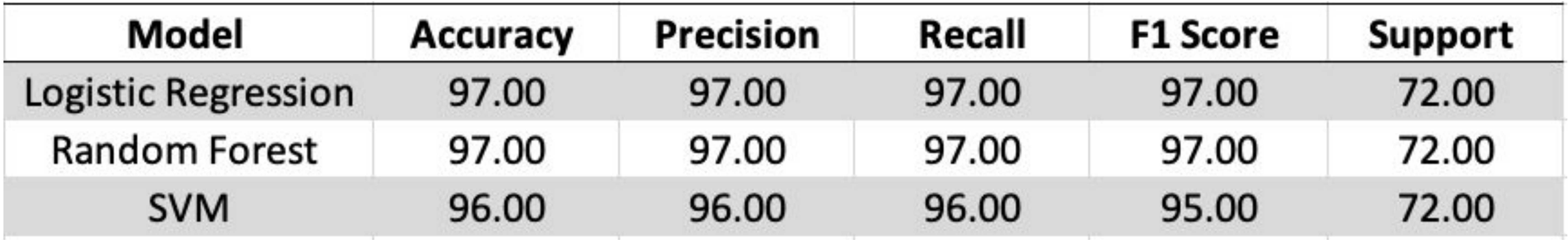
As a consulting service , we want to help automakers of fully electric vehicles (EV) maximize their EV marketing budgets.

What if automakers could target marketing to just consumers that would be the the most likely to buy an EV ?

The solution ?

- **Create a model that can best predict who will purchase an EV**

Model Results



Model	Accuracy	Precision	Recall	F1 Score	Support
Logistic Regression	97.00	97.00	97.00	97.00	72.00
Random Forest	97.00	97.00	97.00	97.00	72.00
SVM	96.00	96.00	96.00	95.00	72.00

The scores are high and similar across models, likely due to the imbalanced nature of the dataset.

Evaluation

Logistic Regression

	Actual Positive	Actual Negative
Predicted Positive	64	0
Predicted Negative	2	6

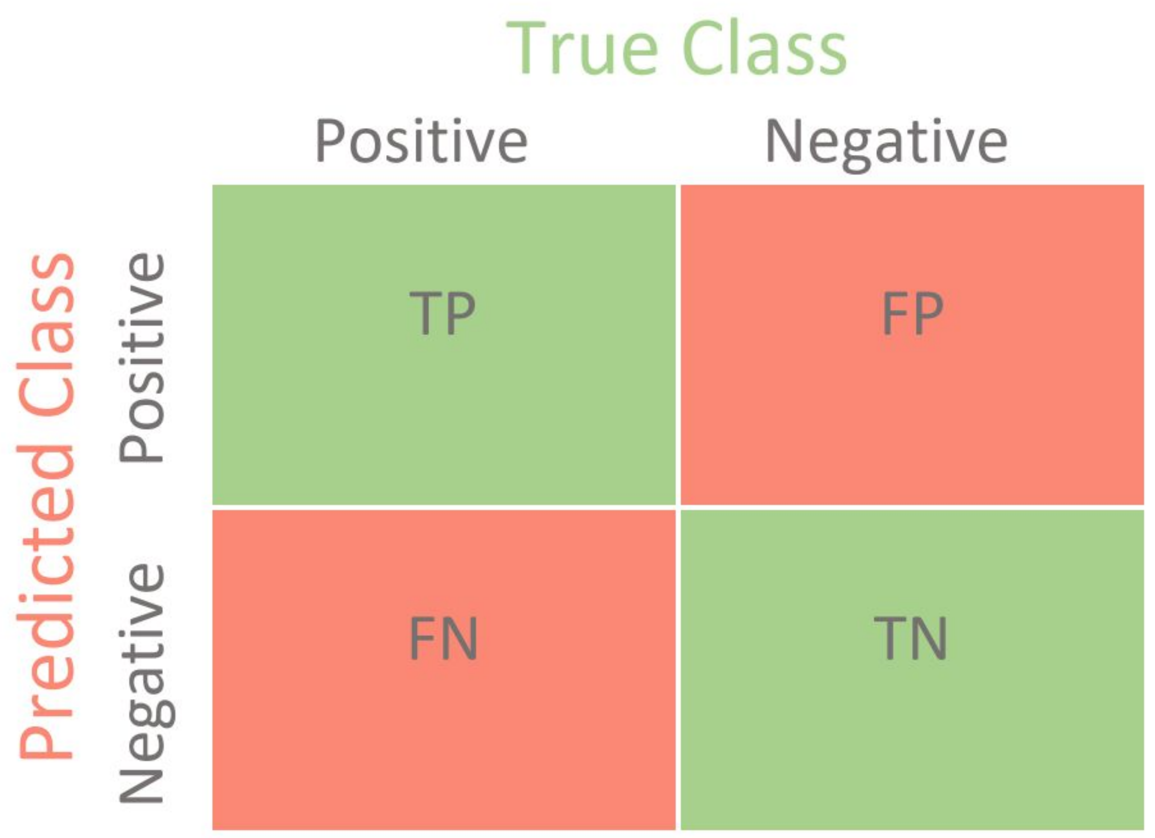
Support Vector Machines

	Actual Positive	Actual Negative
Predicted Positive	64	0
Predicted Negative	2	6

Random Forest

	Actual Positive	Actual Negative
Predicted Positive	64	0
Predicted Negative	2	6

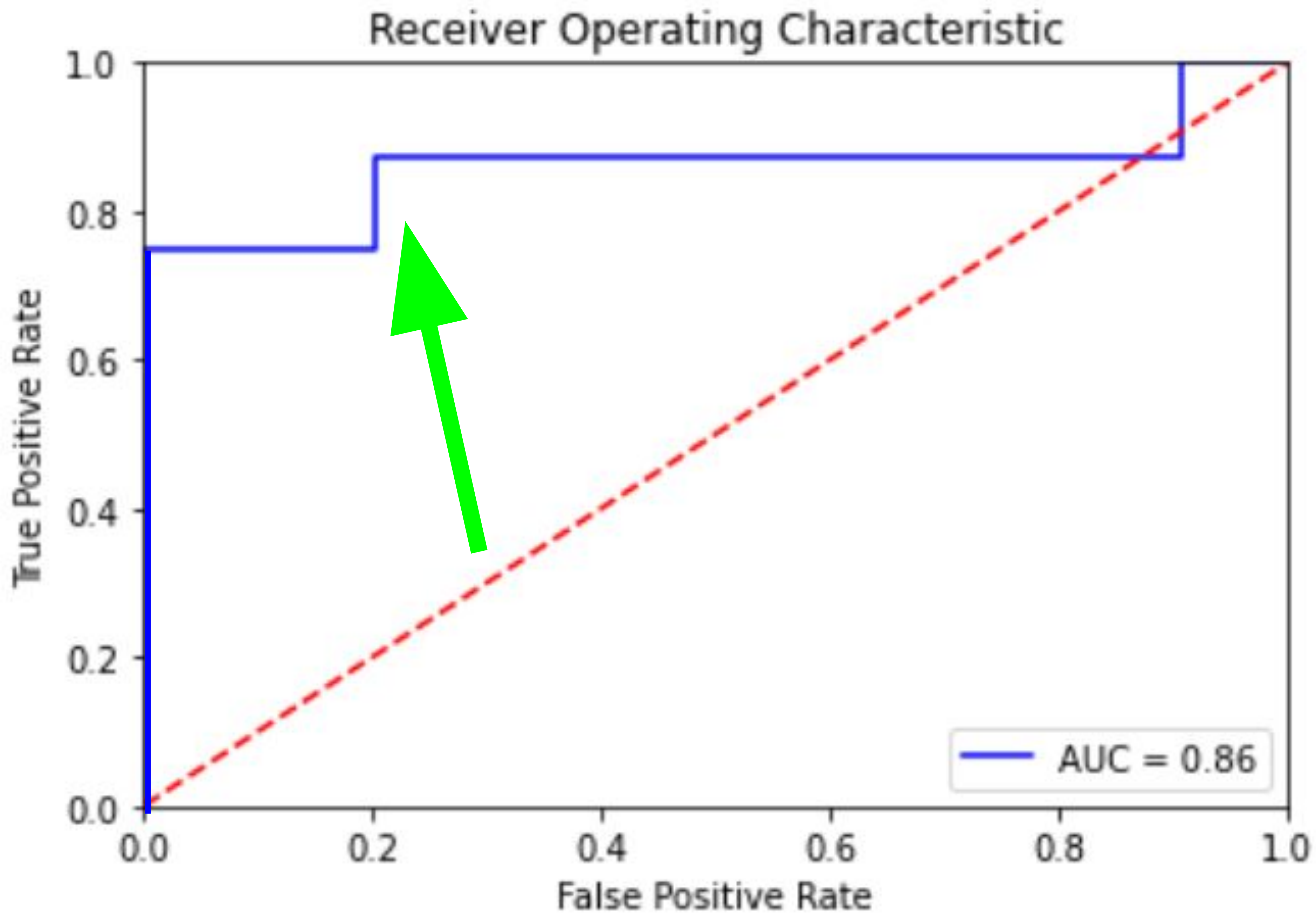
Confusion Matrix : 70% Training 30% Testing



		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

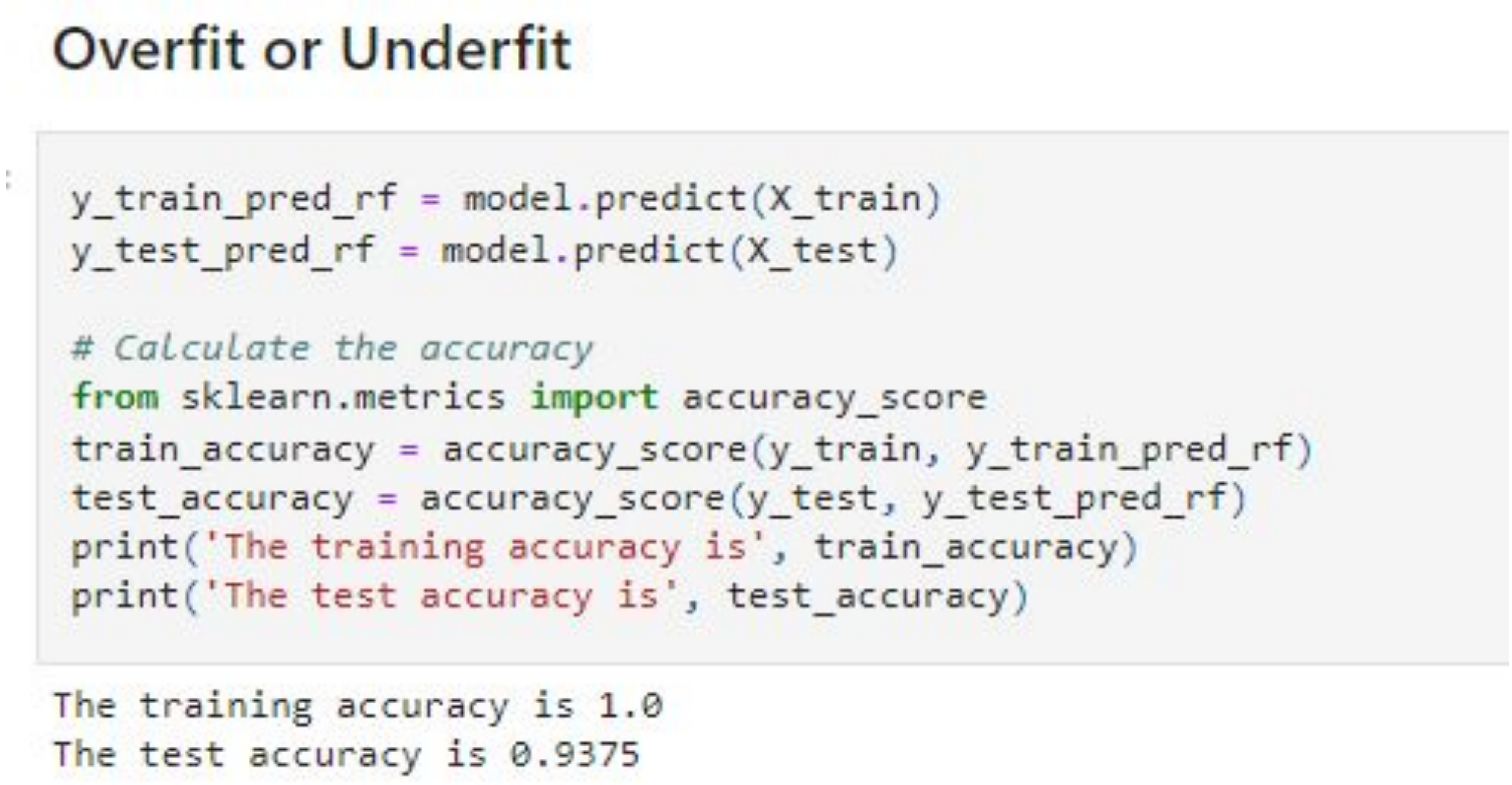
The cost of a False Positive (FP) is miniscule in front of the profit of a True Positive (TP).

ROC Curve



Our broader goal is to maximize True Positives *and* minimize False Positives by picking best tradeoff

Concerns



Overfit or Underfit

```
y_train_pred_rf = model.predict(X_train)
y_test_pred_rf = model.predict(X_test)

# Calculate the accuracy
from sklearn.metrics import accuracy_score
train_accuracy = accuracy_score(y_train, y_train_pred_rf)
test_accuracy = accuracy_score(y_test, y_test_pred_rf)
print('The training accuracy is', train_accuracy)
print('The test accuracy is', test_accuracy)
```

```
The training accuracy is 1.0
The test accuracy is 0.9375
```

- Utilized a 5-fold cross validation technique to attempt to prevent overfitting
- However, our training sets' accuracy is still higher than our test sets' accuracy with a discrepancy of 6%
- Overall, select a larger data set so we can train our model with more data

Business Implications & Deployment

- Overall, select a larger data set so we can train our model with more data

- Overfitting will eventually dissipate as our business grows and we have the financial capital to purchase more data

- Deliver and deploy our results as a consulting service

- Use our model to transform our client's automobile business and give them the competitive edge by helping them accurately market their products and services to the right consumers

Thank You

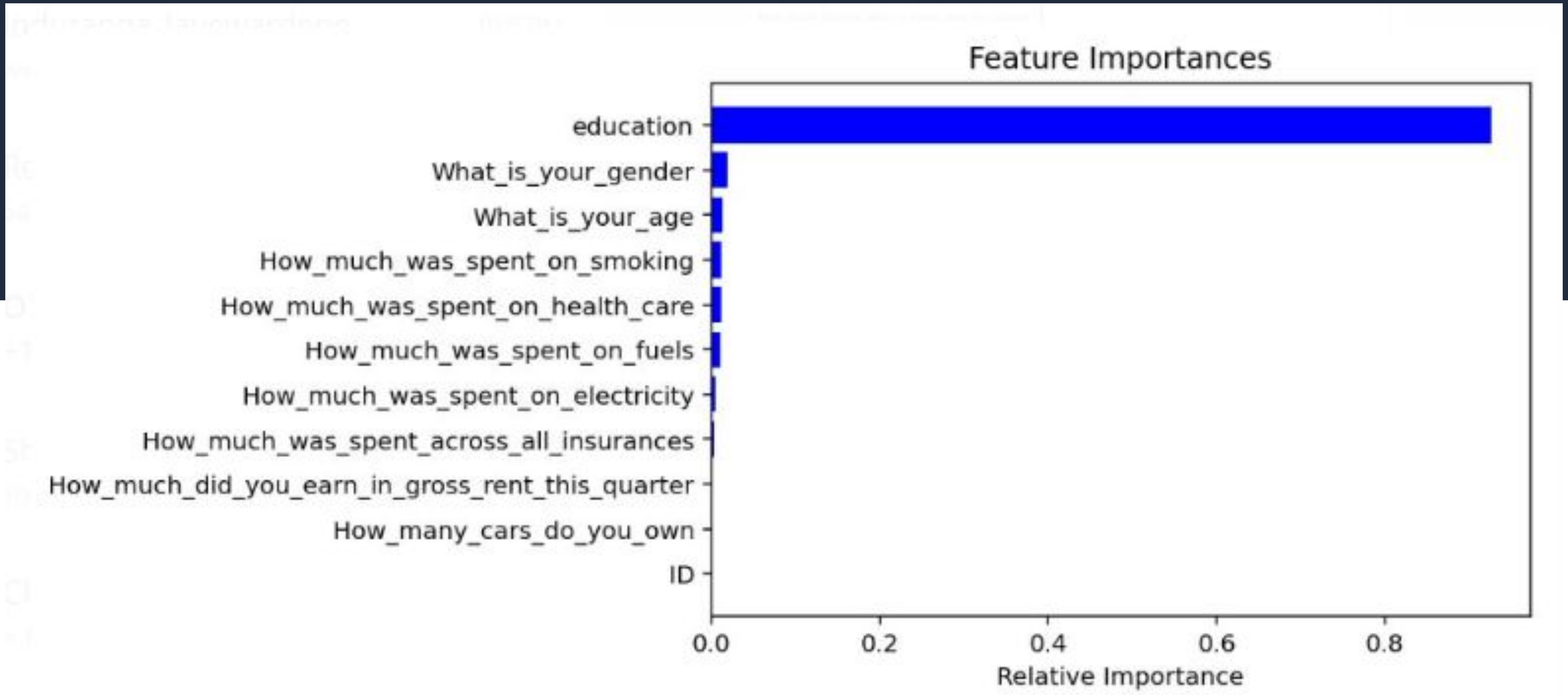
Any Questions ?

Thank You

APPENDIX

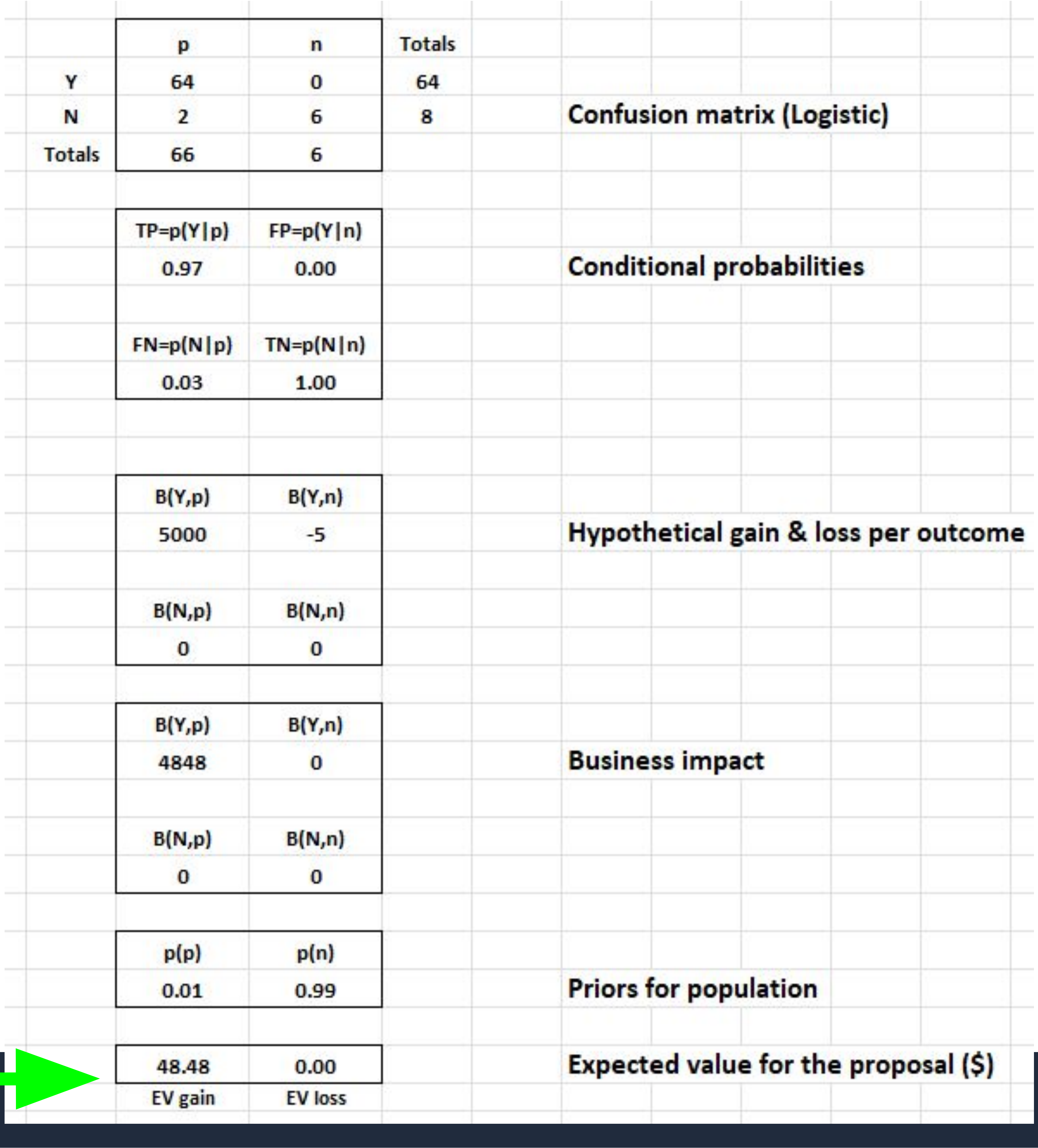
Source Code : https://github.com/cpasean/MSBA/blob/main/BUS9430_Moilestone3.ipynb

FEATURE SELECTION



Source Code : https://github.com/cpasean/MSBA/blob/main/BUS9430_Moilestone3.ipynb

Business Impact & Expected Value



	p	n	Totals
Y	64	0	64
N	2	6	8
Totals	66	6	

Confusion matrix (Logistic)

TP=p(Y|p)	FP=p(Y|n)
0.97	0.00
FN=p(N|p)	TN=p(N|n)
1.00	1.00

Conditional probabilities

B(Y,p)	B(Y,n)
5000	-5
B(N,p)	B(N,n)
0	0

Hypothetical gain & loss per outcome

B(Y,p)	B(Y,n)
4848	0
B(N,p)	B(N,n)
0	0

Business impact

p(p)	p(n)
0.01	0.99

Priors for population

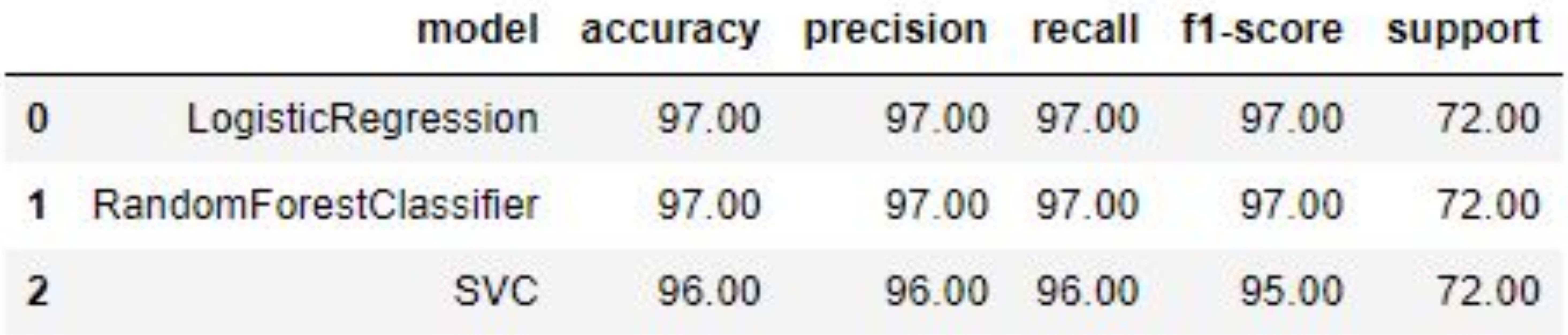
48.48	0.00
EV gain	EV loss

Expected value for the proposal ($)

The proposal has a positive expected value outcome of 48.48

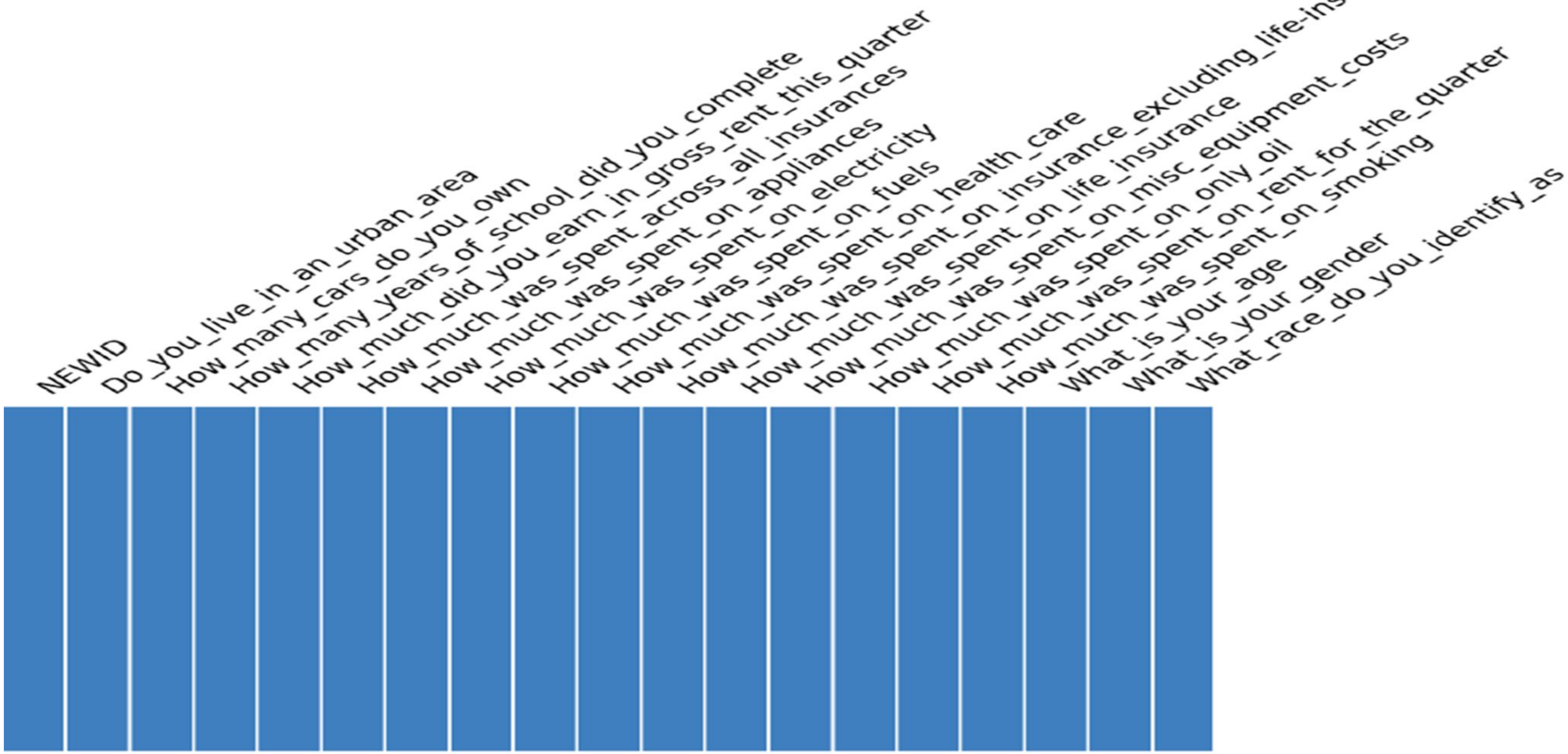
Model

Results**



	model	accuracy	precision	recall	f1-score	support
0	LogisticRegression	97.00	97.00	97.00	97.00	72.00
1	RandomForestClassifier	97.00	97.00	97.00	97.00	72.00
2	SVC	96.00	96.00	96.00	95.00	72.00

Data Columns Tested


NEWID
Do_you_live_in_an_urban_area
How_many_cars_do_you_own
How_many_years_of_school_did_you_complete
How_much_did_you_earn_in_gross_rent_this_quarter
How_much_was_spent_across_all_insurances
How_much_was_spent_on_appliances
How_much_was_spent_on_electricity
How_much_was_spent_on_fuels
How_much_was_spent_on_health_care
How_much_was_spent_on_insurance_excluding_life-insurance
How_much_was_spent_on_life_insurance
How_much_was_spent_on_misc_equipment_costs
How_much_was_spent_on_only_oil
How_much_was_spent_on_rent_for_the_quarter
How_much_was_spent_on_smoking
What_is_your_age
What_is_your_gender
What_race_do_you_identify_as