AFL Sports Fans' Love of Their Team and Its Impact

MKTG6010

Yolanda Kurniarto 480447849

Suya Chen 490247372

Lei Zhang 510221814

Introduction

This project evaluated survey data from Sydney Swans' fans. This project explores how their love for the team affects Sydney Swans' club business performance. For this project, our team will use four main ML techniques to identify patterns in the data: Logit Model, Random Forests, Gradient Boosting, and Support Vector Machines (SVM). Results are then translated into actionable insights and recommendations to improve Sydney Swans' brand performance in brand loyalty, customer retention, and game attendance.

Problem Definition

Based on the goal, the scope for the project is then narrowed down and broken down into three main questions:

- Q1: Does Sydney Swans' on-field performance impact game attendance?
- Q2: What influences Sydney Swans' fans to become loyal supporters?
- Q3: How can Sydney Swans improve their season pass renewal intention?

The dataset used for this project is a subset of a larger dataset as provided by the course, which includes the following features and respondent variables (Table 1):

Question	Dependent Variables	Features Variables
Q1: How does Sydney Swans on-field performance impact game attendance?	Q1_7: Which of the following home games did you attend last year?	Q6 and Q7
Q2: What influences Sydney Swans' fans to become loyal supporters?	Q2_11: I am a long time Swans supporter	Q2_1 until Q2_8, Q5_5 until Q5_8, Q10 (exc. Q10d)
Q3: How can Sydney Swans improve their season pass renewal intention?	Q14_6: When my Swans membership renewal comes, I do not have to consider whether or not to renew; renewal is automatic.	Q5_1, Q5_2, Q9_1, Q9_2, "home_game_sum"

Table 1

For the analysis, the dependent variables are manipulated to reflect binary variables. The sum of all game attendance is calculated for the first dependent variable. Bin 1 notes if the sum of Q1.7 is more significant than a specific level, else 0. In addition, for the brand loyalty variable, a rating of >= 6 is named "loyal". The last variable, called "renewal", is derived from a dummy variable based on answers with a rating of 6 or more.

Q1: Does the on-field performance impact game attendance?

This question requires data from Q1.7, Q6, and Q7. Q6 represents the participants' individual opinions of the importance of each aspect in Q7, and Q7 represents the participants' evaluation of the Sydney Swans' performance. After considering the characteristics of each question, the binary model is split into three parts. The first model includes all columns from Q6 and Q7. The second model includes Q7 only. The last one is the five major questions of Q7 weighted by the score of Q6. The model evaluation uses binary and multiclass models.

Binary Model

Three attempts of the binary model were conducted for model and feature evaluation. The comparison of the three models is presented in Table 2.

Table 2

Attempt	Model	Error rate	Precision	Sensitivity	Specificity	AUC	Cross-Entropy
1	**Logit**	0.444444	0.546218	0.822785	0.270270	0.604858	0.694465
	RF	0.483660	0.517730	0.924051	0.081081	0.601437	0.689445
	XGBoost	0.522876	0.494737	0.594937	0.351351	0.493158	1.412923
	SVM	0.509804	0.503704	0.860759	0.094595	0.620253	
2	**Logit**	0.424837	0.554688	0.898734	0.229730	0.612641	0.683489
	RF	0.483660	0.517730	0.924051	0.081081	0.651642	0.678135
	XGBoost	0.411765	0.580000	0.734177	0.432432	0.619141	0.983438
	SVM	0.483660	0.516340	1.000000	0.000000	0.613325	
3	**Logit**	0.490196	0.513514	0.962025	0.027027	0.533014	0.699885
	RF	0.535948	0.488889	0.835443	0.067568	0.503079	0.719077
	XGBoost	0.490196	0.523810	0.556962	0.459459	0.537975	1.136315
	SVM	0.483660	0.516340	1.000000	0.000000	0.474855	

The first model attempts to predict individual attendance according to the response to all the questions in section 3. The metrics (Table 2) indicates that the Logit model slightly outperforms the others. According to the feature importance (*appendix 1*), responses to questions 7.1, 7.4, and 7.2 are significant indicators for the dependent variable.

The second model attempts to predict individual attendance similarly to the first but only includes responses from Q7. The metrics (Table 2) indicate that the Gradient Boost model outperforms the others. In addition, according to the feature importance (Figure 2), responses to questions 7.4, 7.22, and 7.3 are significant indicators for the dependent variable.

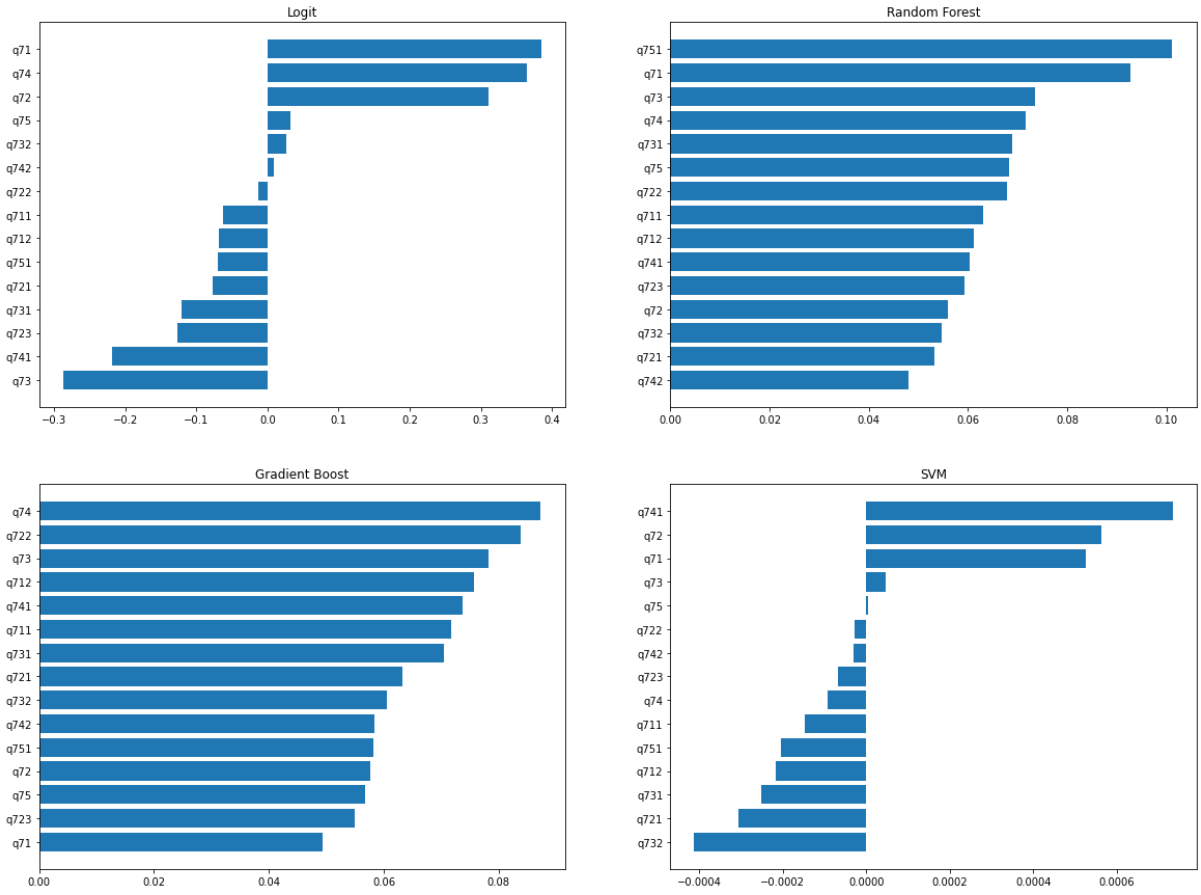


Figure 1

The third model attempts to predict individual attendance based on the responses to main questions in Q7 weighted by the scores in Q6. The metrics shown above indicate that the Gradient Boost model outperforms the others. Meanwhile, according to the feature importance (*appendix 2*), responses to questions 7.4 and 7.3 are significant indicators for the dependent variable.

Apart from the models above, a multiclass model would be somewhat helpful when the dependent variable is an integer between 0 and 10. According to the feature importance (*appendix 3*), the response to question 7.4 is a significant indicator for the dependent variable.

Recommendations

In conclusion, the performance of the Sydney Swans does impact match day attendance.
The metrics of each model indicate that the second binary model, Gradient Boost on Q7, outperforms the others.

Responses to questions 7.4, 7.22, and 7.3 are considered significant indicators of attendance. The Sydney Swans shall work on these aspects to increase their match-day revenue, i.e., they shall endeavour to play as a team, keep calm and collected under high pressure, and care about their fans. In general, the players shall attempt more team play on the pitch. When opponents press forward, they must remain calm. Also, they could try to entertain the crowd, when possible, to demonstrate their care for the fans.

Q2: What influences Sydney Swans' fans to become loyal supporters?

In investigating the factors that may influence Sydney Swans' fans to become loyal supporters, Q2.11 I am a long time Swans supporter was chosen as the target variable. This variable measures whether respondents regard themselves as loyal fans of the Swans. Initially, the variable is quantified by a Likert 7-scale in which the respondents choose whether this statement describes them well. The target variable was then manipulated to reflect a binary variable by converting it to a dummy variable “loyal”.

Model and Feature Evaluation

Four different models are initially trained with independent variables from Q2_1 until Q2_8 and Q5_5 until Q5_8 as independent variables. The variables from Q2 are chosen from the survey because we wish to evaluate whether social factors, which include reference groups such as family and friends, influence the attitude and opinions of respondents (Lautiainen 2015) towards Sydney Swans. Additionally, a sense of belonging, which may develop brand

loyalty and long-term marketing impact (Bateman 2021), is also investigated from Q5_5 until Q5_8 from the survey.

Based on incorporating these independent variables, the SVM model, which has the lowest error rate, is then chosen as the benchmark to measure the model's performance and features. Next, the process continues to feature selection using the forward selection strategy (Jovic, Brkic, Bogunovic 2015); more features are added to the set to evaluate whether the model performance will increase. It can be seen from comparing the error rate that the model performance increase, although not significant, as extra independent variables from Q10 (excluding Q10d) are added. A summary of the model result can be seen in Table 3:

Table 3

Attempt	Model	Error Rate	Precision	Sensitivity	Specificity	AUC	Cross-Entropy
1	**Logistic**	0.2	1.0	0.6	1.0	0.86	0.570
	Random Forest	0.2	1.0	0.6	1.0	0.97	0.414
	XGB	0.3	0.75	0.6	0.8	0.9	0.553
	SVM	0.15	1.0	0.7	1.0	0.91	-
2	**Logistic**	0.15	1.0	0.7	1.0	0.9	0.521
	Random Forest	0.2	1.0	0.6	1.0	0.96	0.428
	XGB	0.15	0.89	0.8	0.9	0.91	0.503
	SVM	0.1	1.0	0.8	1.0	0.91	-

The result shows that the error rate for SVM, Logistic, and XGB models decreased, indicating that the models perform better with higher accuracy. Since SVM is the top performer out of all models, the SVM model with 16 independent variables listed in Figure 2 is chosen as the final model.

Recommendation

The features that are included in the final model comparison are social and environmental factors (Q2_1 until Q2_8), a sense of belonging (Q5_5 until Q5_8), and feelings towards Sydney Swans (Q10, etc. Q10d).

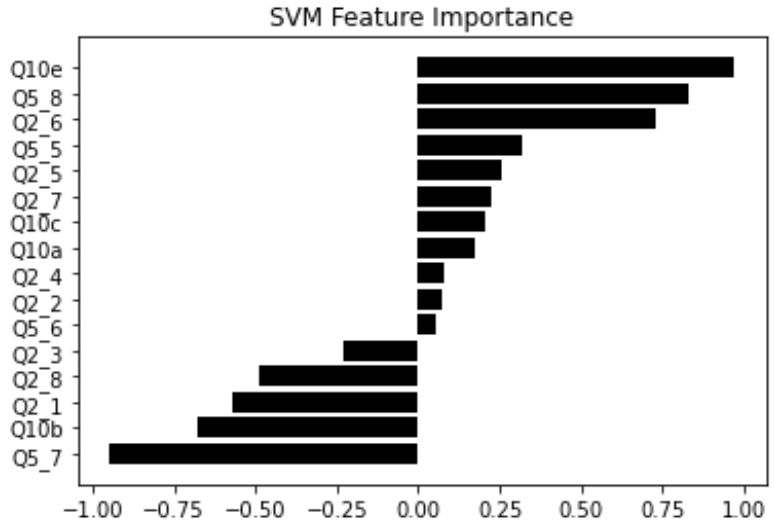


Figure 2

Focusing on the top three most critical features (Figure 5): Q10e, Q5_7, and Q5_8, respondents feel a sense of belonging in which they consider Swans' victory as their own (Q5_7 and Q5_8). This sense of belonging is supported by their confidence (Q10e) in Sydney Swans, inspiring them to become loyal supporters of the team.

As such, it is recommended that stores leverage this pride as a Swans' community by providing exclusive merchandise and access for members, such as a private lounge or meet and greet event. This sense of belonging can be further amplified if members gather for social events other than game matches to leverage their shared affections and develop loyalty to the team.

Q3: How can Sydney Swans improve their season pass renewal intention?

In investigating this question, the independent variables are chosen as follows. Firstly, being a long-term member of a team requires fans to have sufficient love and loyalty to the team, so we select Q5_1 and Q5_2 as the relevant variables. Secondly, the management of the club, especially the part directly related to the fans, will also be a priority for us to include. We, therefore, have Q9_1 and Q9_2 in the selection of independent variables. Again, the total number of games attended will be an essential indicator of affordability, willingness to spend,

love and loyalty. We, therefore, include “home_game_sum” in the selection of the independent variables.

Model and Feature Evaluation

Table 4

Model	Error rate	Precision	Sensitivity	Specificity	AUC	Cross-Entropy
Logit model	0.262195	0.8125	0.757282	0.704918	0.795082	0.536252
Random Forest	0.237805	0.840426	0.76699	0.754098	0.812749	0.523759
XG Boosting	0.335366	0.772727	0.660194	0.672131	0.765319	0.957775
SVM	0.256098	0.814433	0.76699	0.704918	0.796355	-

From Table 4, the random forest model performs best and as such was chosen as the final. The error rate of random forest, 0.2378, is the lowest among the four models, indicating the highest accuracy of all case classifications. The AUC is the highest out of all, which means random forest performs well in telling positive and negative cases.

Recommendation

From the feature importance of random forest (Figure 3), the most important feature is q52 (Supporting the Swan is an important part of my life), followed by q51 (It’s important to me to maintain the ties with Sydney Swan), and lastly, q912 (Try to get members finals tickets), which suggest the key point to improve the season pass renew rate is to focus on the fan with strong loyalty and love, and Sydney Swans could design a bonus system to encourage this group of fans to renew their pass, for example, the longer they hold the pass, the higher the chance they could get the limited finals ticket offered by Sydney Swans.

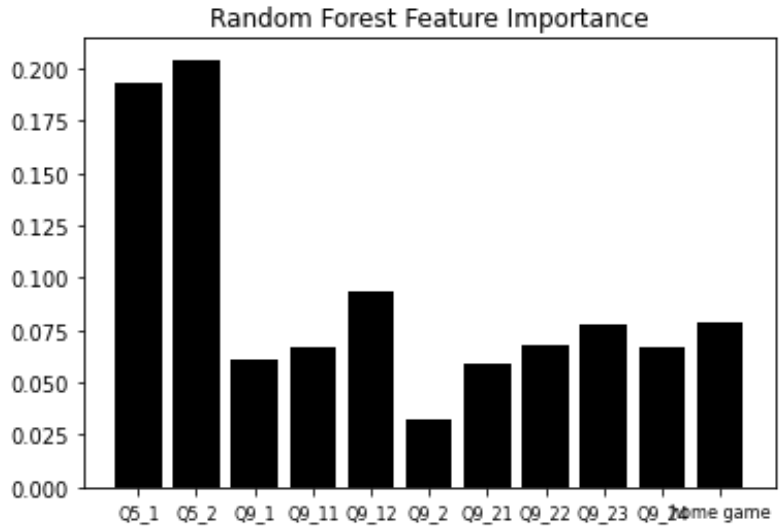


Figure 3

Limitations and Future Directions

The findings of this study are not without limitations. The primary caveat is that the initial feature selection method mainly relies on the researcher's judgment and without consideration of the whole feature variables available in the dataset. It is advised that future studies incorporate more sophisticated feature selection methods such as filtering based on information gain, chi-square, linear forward selection (Jovic, Brkic, & Bogunovic 2015) or others.

Reference List

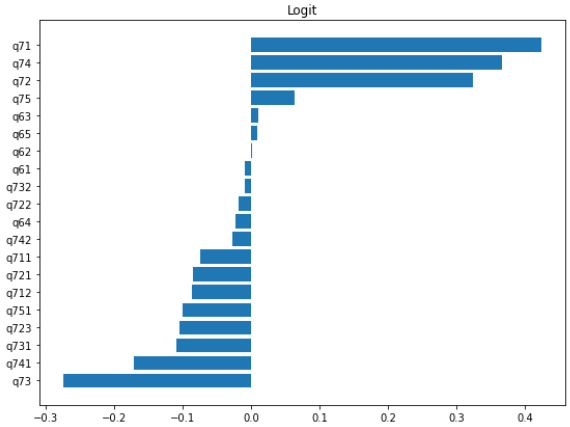
Bateman, A 2021, 'Why belonging is the new customer loyalty and three ways to cultivate it' *Forbes*, 12 May, https://www.forbes.com/sites/forbesbusinesscouncil/2021/05/12/why-belonging-is-the-new-customer-loyalty-and-three-ways-to-cultivate-it/?sh=307e47a71185

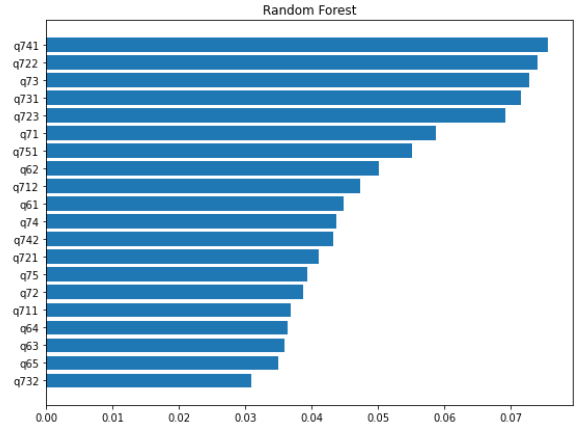
Jovic, A, Brkic, K, & Bogunovic, N 2015, 'A review of feature selection methods with applications' *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1200–1205. https://doi.org/10.1109/MIPRO.2015.7160458

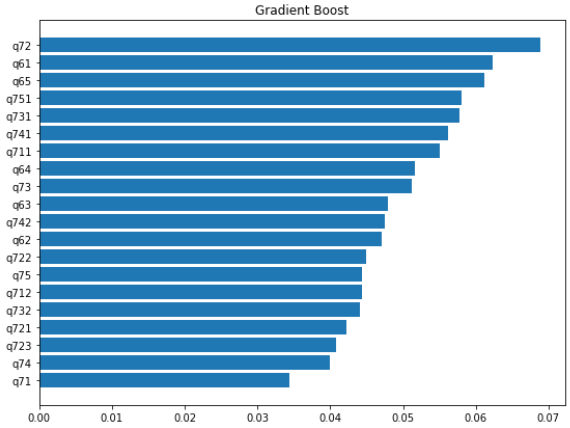
Lautiainen, T 2015, 'Factors affecting consumers' buying decision in the selection of a coffee brand' Bachelor's thesis, Saimaa University of Applied Sciences.

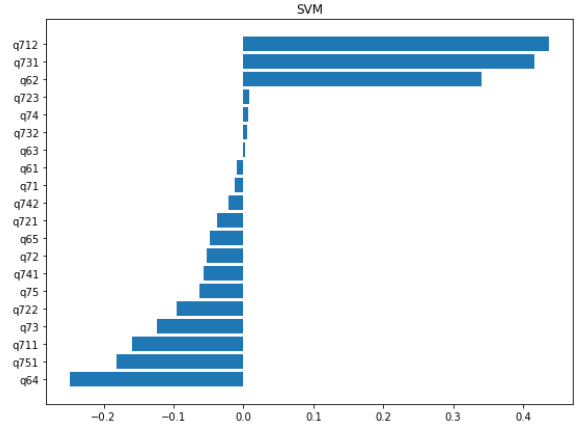
Appendix

Appendix 1

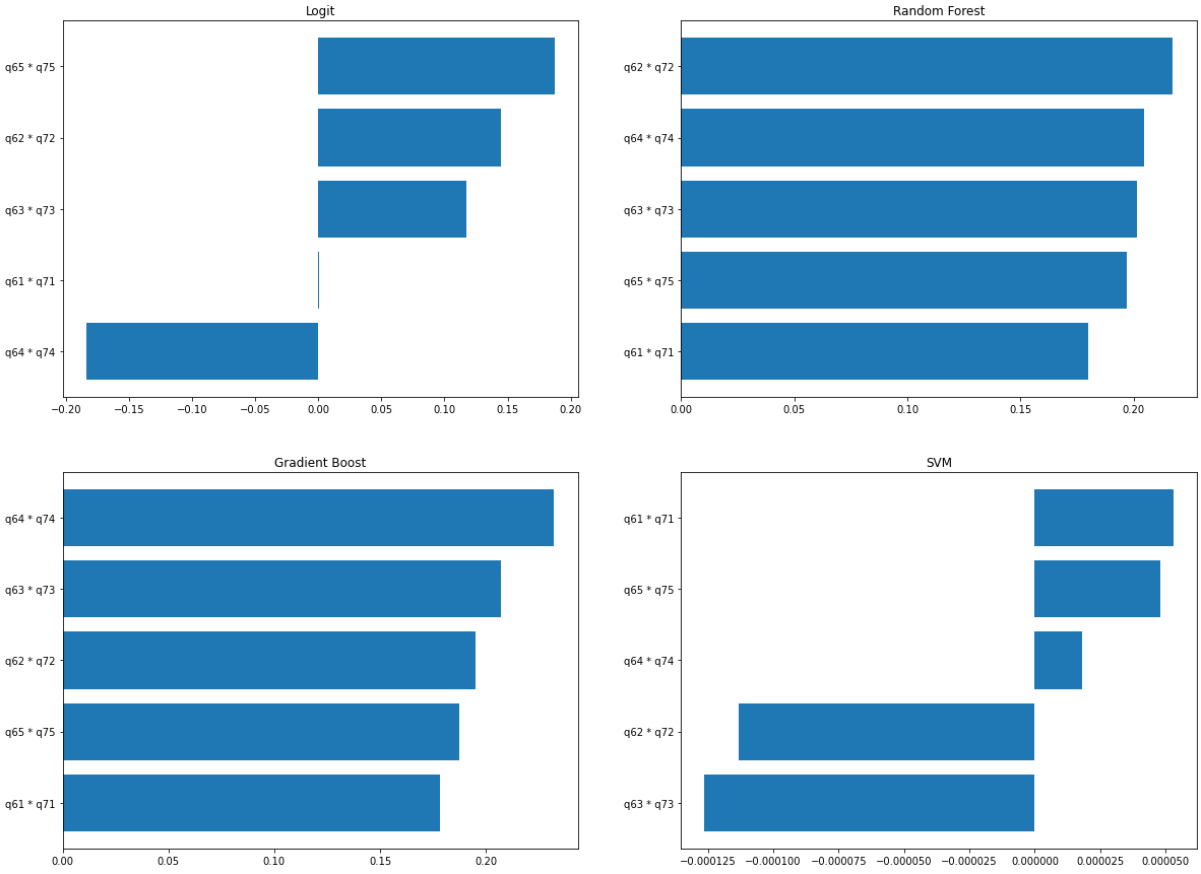




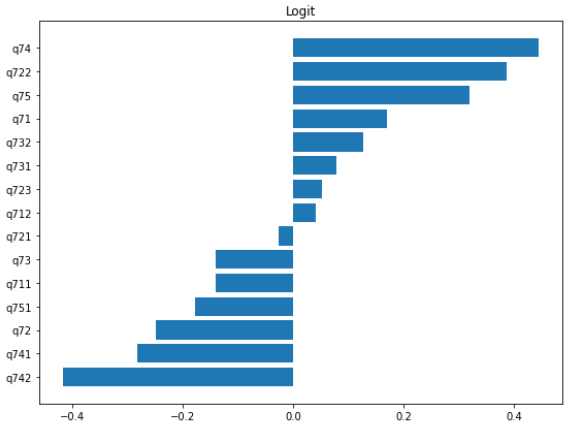


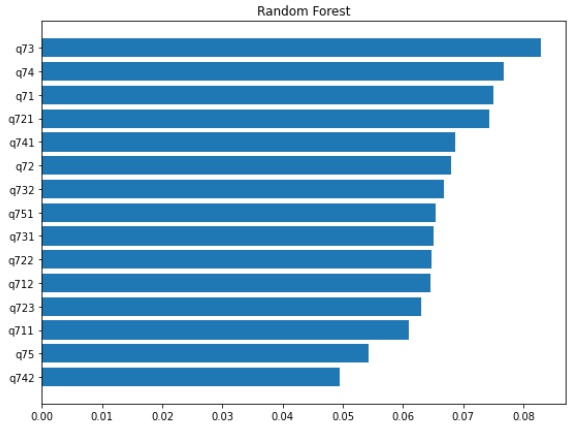


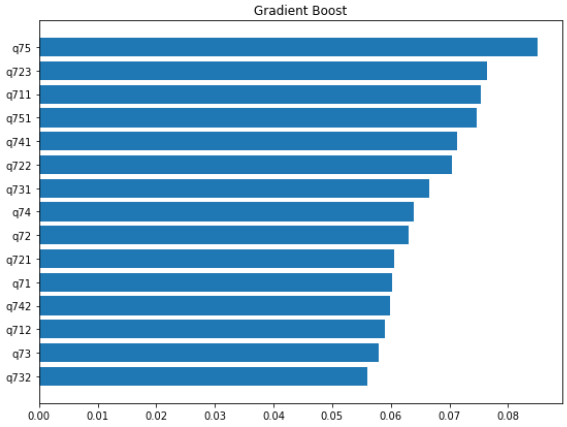
Appendix 2

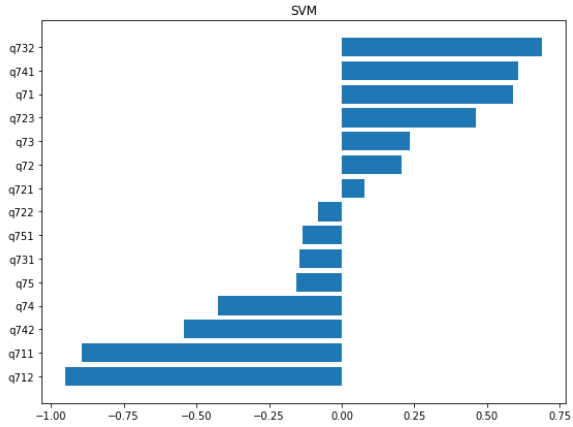

Logit
q65 * q75
q62 * q72
q63 * q73
q61 * q71
q64 * q74
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
Random Forest
q62 * q72
q64 * q74
q63 * q73
q65 * q75
q61 * q71
0.00
0.05
0.10
0.15
0.20
Gradient Boost
q64 * q74
q63 * q73
q62 * q72
q65 * q75
q61 * q71
0.00
0.05
0.10
0.15
0.20
SVM
q61 * q71
q65 * q75
q64 * q74
q62 * q72
q63 * q73
−0.000125
−0.000100
−0.000075
−0.000050
−0.000025
0.000000
0.000025
0.000050

Appendix 3


Logit
q74
q722
q75
q71
q732
q731
q723
q712
q721
q73
q711
q751
q72
q741
q742
−0.4
−0.2
0.0
0.2
0.4


Random Forest
q73
q74
q71
q721
q741
q72
q732
q751
q731
q722
q712
q723
q711
q75
q742
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08


Gradient Boost
q75
q723
q711
q751
q741
q722
q731
q74
q72
q721
q71
q742
q712
q73
q732
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08


SVM
q732
q741
q71
q723
q73
q72
q721
q722
q751
q731
q75
q74
q742
q711
q712
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75