Bank Customer Churn

Kylinn Kraemer

Business Case

Link to dataset

Our key stakeholders are the **C-suite of a large banking firm** with a vested interest in reducing customer churn.

This data will undergo preprocessing and analysis to identify methods to **reduce customer churn**.

Objective:

To minimize customer churn

Data: Summary

10,000	20%	10
rows of **customer data** were analyzed	of customers **churned**	**features** were analyzed to predict churn

Market average churn is 15%

Data Dictionary

Categorical/Binary Variables

country: the customer's country of residence

gender: the gender of the customer

credit card: if the customer has a credit card

active_member: if the customer is an active member of the bank

products_number: the bank product the customer used

churn: if the customer has churned

Quantitative Variables

credit_score: the credit score of the customer

age: the age of the customer

tenure: how many years the customer has had an account with the bank

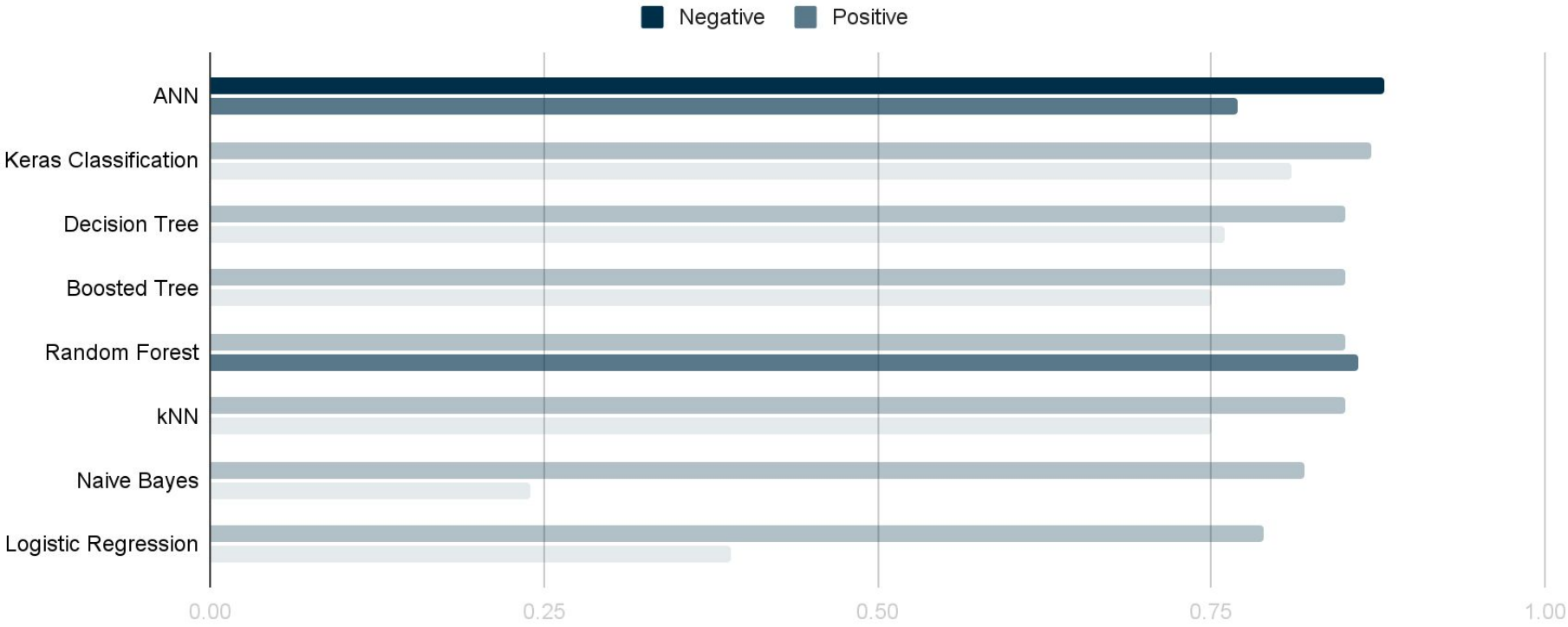
balance: account balance

estimated_salary: estimated salary of the customer

Methodology

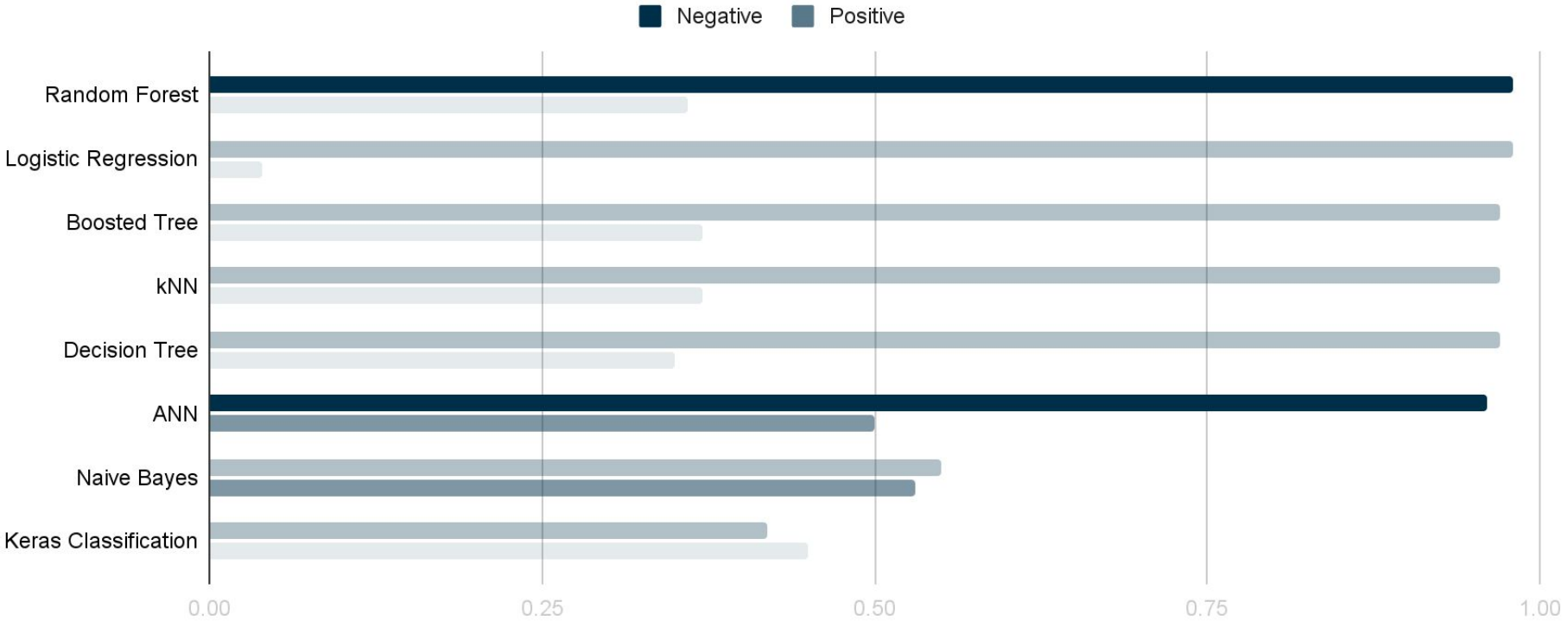
We employed a variety of machine learning techniques to develop a model capable of **predicting customer churn**.

Investigative analysis techniques were used to identify variables with a high **impact on customer churn**.

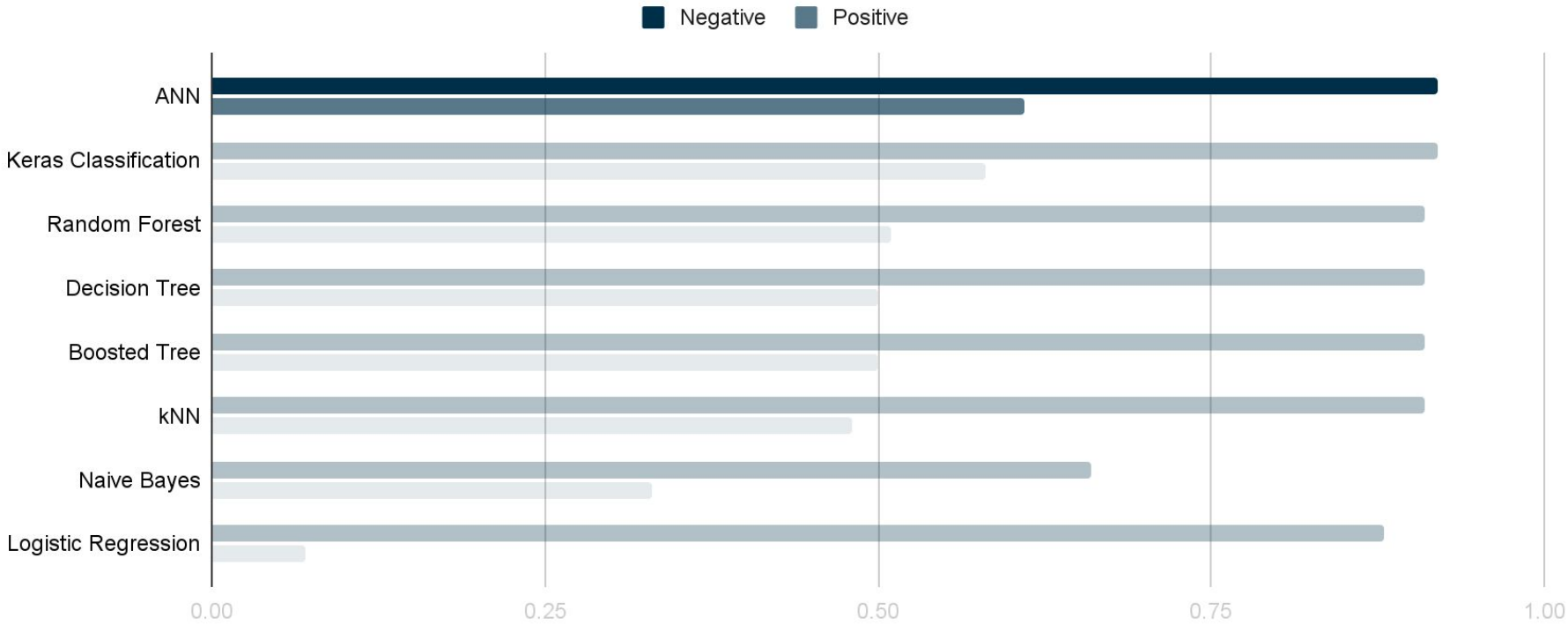


The ANN model has the highest positive precision, but noticeably lower negative precision than the best performing model - the Random Forest.

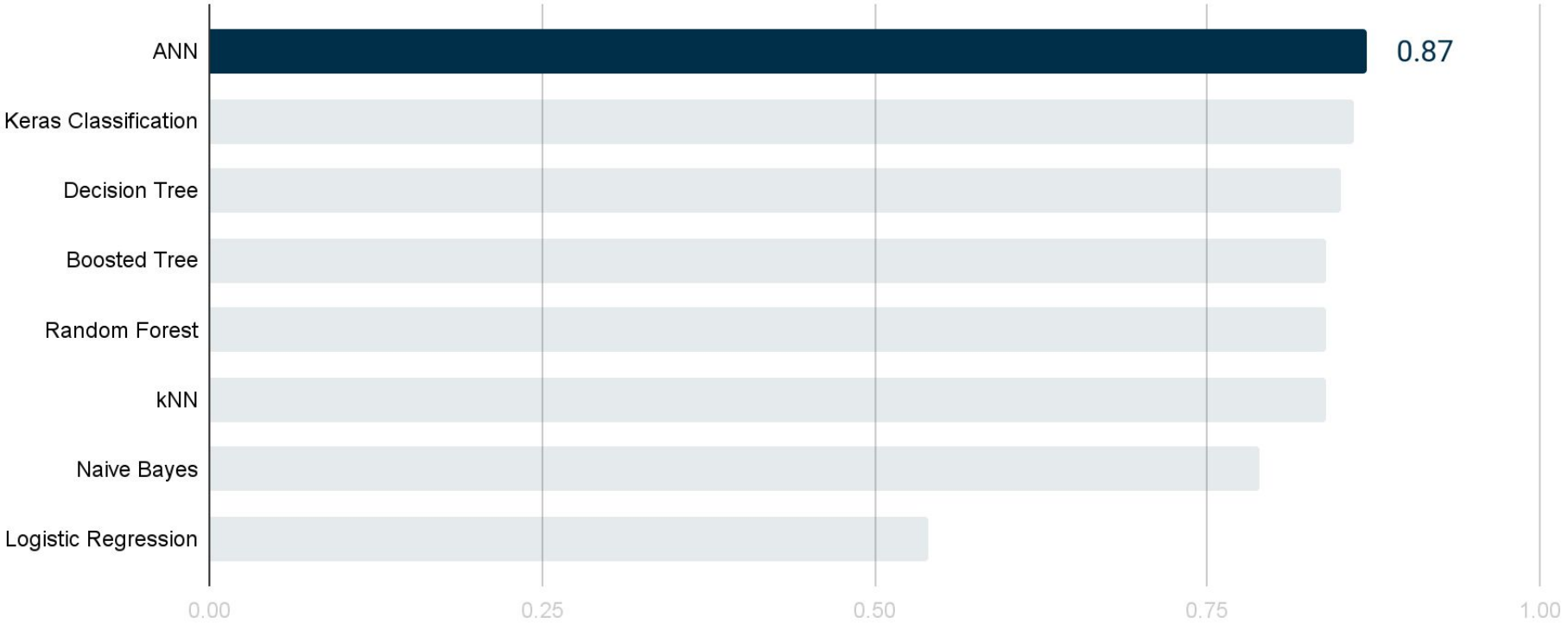
The Random Forest has the highest positive recall and the Naive Bayes has the highest negative recall.



However, the ANN model has a comparable positive and negative recall to the best performing models.

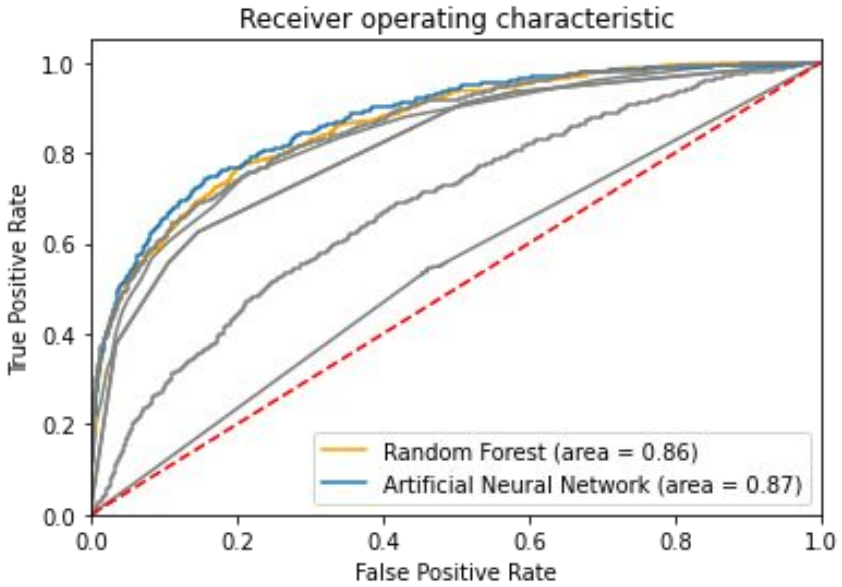


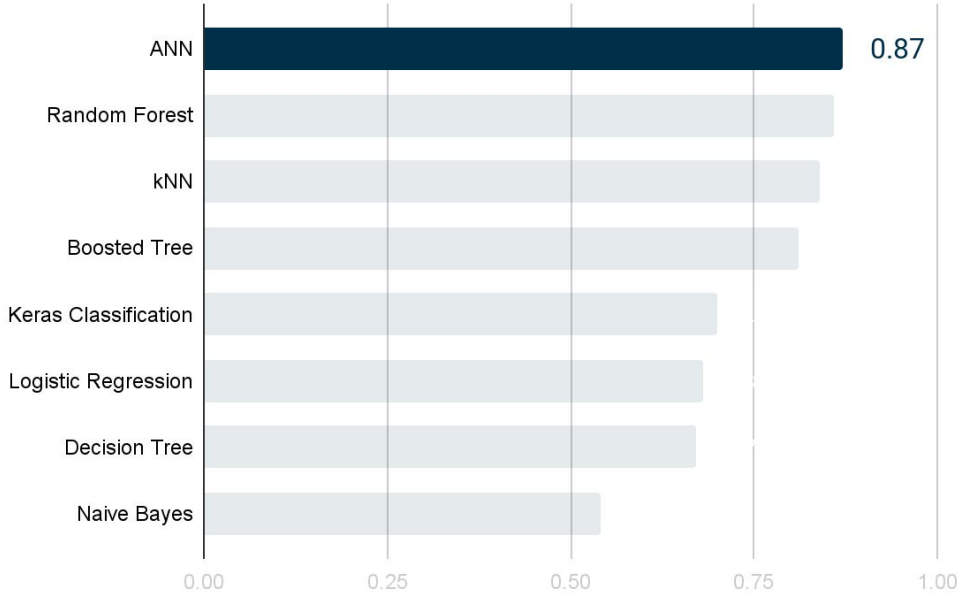
The ANN model has the highest F1 score in terms of identifying negatives and positives.



The ANN model also has the highest accuracy.

The ANN Model has the highest ROC AUC!

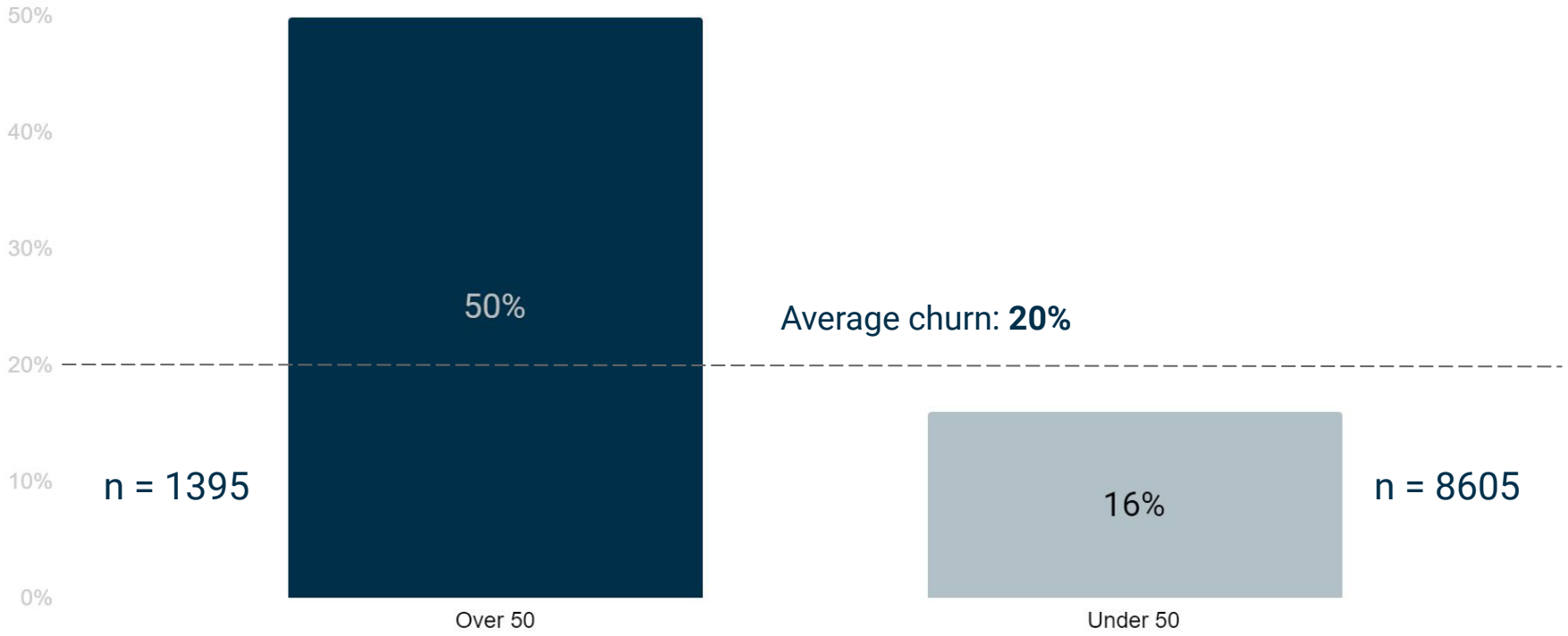




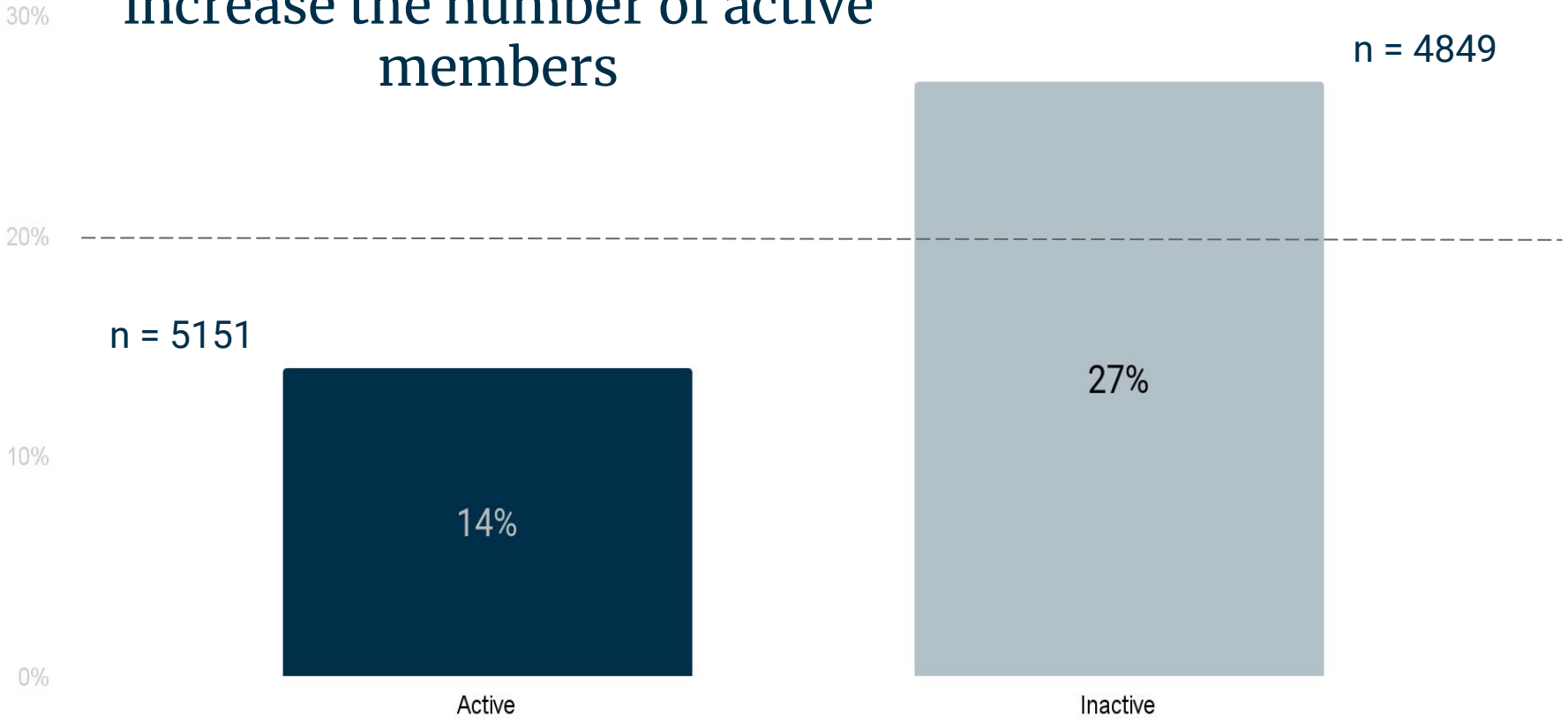
	F1 (positive)	F1 (negative)	Precision (positive)	Precision (negative)	Recall (positive)	Recall (negative)	Accuracy	ROC AUC
Logistic Regression	0.07	0.88	0.39	0.79	0.04	0.98	0.79	0.68
kNN	0.48	0.91	0.76	0.85	0.35	0.97	0.84	0.84
Decision Tree	0.5	0.91	0.75	0.85	0.37	0.97	0.84	0.67
Random Forest	0.51	0.91	**0.86**	0.85	0.36	**0.98**	0.85	0.86
Boosted Tree	0.5	0.91	0.75	0.85	0.37	0.97	0.5	0.81
Naive Bayes	0.33	0.66	0.24	0.82	**0.53**	0.55	0.54	0.54
ANN	**0.61**	**0.92**	0.77	**0.88**	0.5	0.96	**0.87**	**0.87**
Keras Classification	0.57	0.92	0.8	0.87	0.44	0.97	0.86	0.7

Recommendations

Target marketing towards customers **over 50**



Scale up engagement efforts to increase the number of active members



Because of high churn and low customer usage, we recommend **discontinuing products 3 and 4**