

**Seems familiar?**





# Iranian Telecommunication Churn Prediction

---

Machine Learning



# Why should Iranian Telecom Company care?



## Touchpoints

- Hard to have engagement with customers
- Importance to build unique customer experiences



## Substitutes

- Unemotional utility service
- People have no high brand preference
- Low switching costs



## Competition

- Competing on costs
- Important to lower CAC & maximize CLV

**Great opportunity to improve retention and create value for the company**

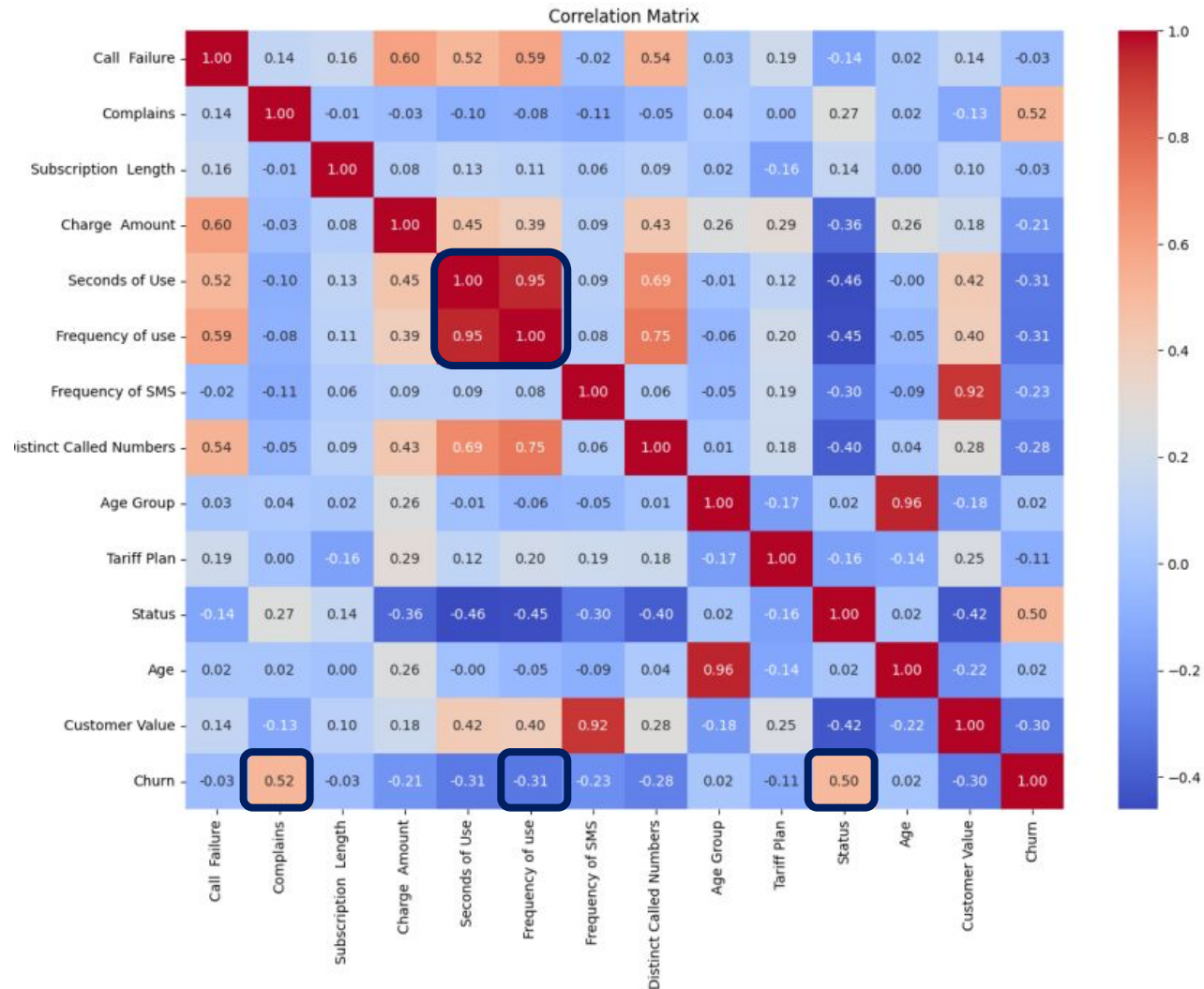
# Introduction to our dataset

With 3,150 instances and 14 variables - 15% churn

2	Call	Failure	Complains	Subscription	Length	Charge	Amount	Seconds of Use	Frequency of use	Frequency of SM	Distinct Called Number	Age Group	Tariff Plan	Status	Age	Customer Value	Churn	proba_0	proba_1	proba_percentile
3	0	0	0	35	0	0	0	0	0	0	0	4	1	2	45	0.0	1	0.0068869807441958	0.993113019255804	100
4	14	1	1	37	0	3295	47	16	18	3	1	2	30	197.68	1	0.0016282807893657	0.9983717192106343	100		
5	8	1	1	11	2	1792	25	7	9	3	1	1	30	100.68	1	0.0015492063492063	0.9984507936507936	100		
6	6	1	1	11	2	1548	22	5	9	3	1	1	30	82.8	1	0.01	0.99	100		
7	0	1	1	32	0	0	0	0	0	3	1	2	30	0.0	1	0.0043755404158386	0.9956244595841615	100		
8	0	0	0	34	0	0	0	0	0	3	1	2	30	0.0	1	0.0067064376982505	0.9932935623017494	100		
9	0	0	0	36	0	0	0	0	0	3	1	2	30	0.0	1	0.0136323905850695	0.9863676094149305	100		
10	0	1	1	34	0	0	0	0	0	3	1	2	30	0.0	1	0.0025488198338781	0.9974511801661218	100		
11	0	0	0	33	0	0	0	0	0	3	1	2	30	0.0	1	0.0071344206039203	0.9928655793960796	100		
12	0	1	1	34	0	0	0	0	0	3	1	2	30	0.0	1	0.0025488198338781	0.9974511801661218	100		
13	0	1	1	32	0	0	0	0	0	2	1	2	25	0.0	1	0.0045552163298157	0.9954447836701843	100		
14	0	0	0	36	0	0	0	0	0	4	1	2	45	0.0	1	0.0084361870934022	0.9915638129065977	100		
15	5	1	1	37	0	2765	37	12	18	2	1	2	25	180.09	1	0.0133596779790408	0.9866403220209592	100		
16	5	1	1	9	1	1733	22	7	7	3	1	1	30	98.2	1	0.01	0.99	100		
17	14	1	1	37	0	3295	47	16	18	3	1	2	30	197.68	1	0.0016282807893657	0.9983717192106343	100		
18	0	1	1	34	0	0	0	0	0	3	1	2	30	0.0	1	0.0025488198338781	0.9974511801661218	100		
19	0	1	1	37	0	503	9	13	7	4	1	2	45	45.3	1	0.0364001690125130	0.963599830987487	99		
20	23	1	1	33	0	955	47	16	17	2	1	2	25	117.09	1	0.0300790744401594	0.9699209255598406	99		
21	12	1	1	34	0	3115	39	7	13	3	1	2	30	154.16	1	0.0233596779790408	0.9766403220209591	99		
22	6	1	1	34	0	2153	39	36	32	3	1	2	30	231.68	1	0.0318104716298345	0.9681895283701656	99		
23	3	0	0	9	1	1488	19	0	7	3	1	1	30	60.28	1	0.0229609029615364	0.9770390970384636	99		
24	0	0	0	33	0	0	0	0	0	4	1	2	45	0.0	1	0.0173149636498656	0.9826850363501343	99		



# Exploratory Data Analysis with Correlation Matrix



## High correlation

- Frequency of Use & Seconds of Use: 0.95
- Frequency of SMS & Customer Value: 0.92

## Substantial correlations with churn

- Complaints & Churn: 0.52
- Status & Churn: 0.50
- Frequency of use & churn : -0.31

# Model Selection - Logistic Regression

Logistic Regression, Support Vector Classifier, and Random Forest



## Strengths

- Binary probability classification
- Explainability



## Weaknesses

- Assumes linear relationships
- Not fitting our dataset

## Model Performance

Metric	Score
Precision	0.7605
Recall	0.4302
F1	0.5496
Added value	13,284.00 €

# Model Selection - Support Vector Classifier

Logistic Regression, Support Vector Classifier, and Random Forest



## Strengths

- Handles high-dimensionality
- Robustness against outliers



## Weaknesses

- Sensitive to overfitting
- Lack of explainability because of black-box approach

## Model Performance

Metric	Score
Precision	0.9375
Recall	0.5378
F1	0.6835
Added value	13,486.50 €

# Model Selection - Random Forest

Logistic Regression, Support Vector Classifier, and Random Forest



## Strengths

- Handling of numerical and categorical variables
- Accounts for non-linear relations
- Interpretability and valid feature importances



## Weaknesses

- Complexity in interpreting individual trees from the ensemble

## Model Performance

Metric	Score
Precision	0.8858
Recall	0.7729
F1	0.8255
Added value	23.233,50 €

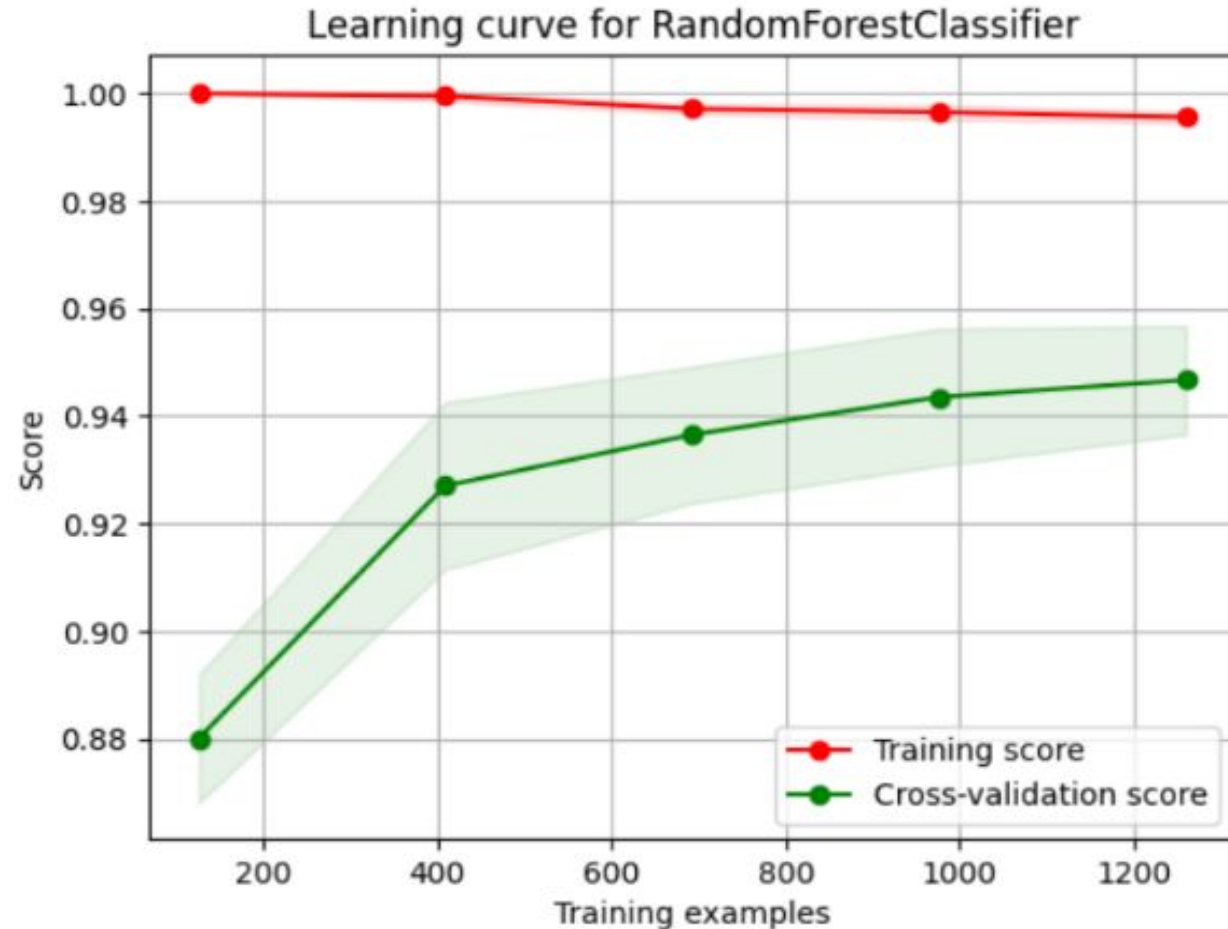


# Model Selection - Side by Side Comparison

Logistic Regression, Support Vector Classifier, and Random Forest

Metric	Logistic Regression	SVC	Random Forest
Precision	0.7605	0.9375	0.8858
Recall	0.4302	0.5378	0.7729
F1	0.5496	0.6835	0.8255
Added value	13,284.00 €	13,486.50 €	23.233,50 €

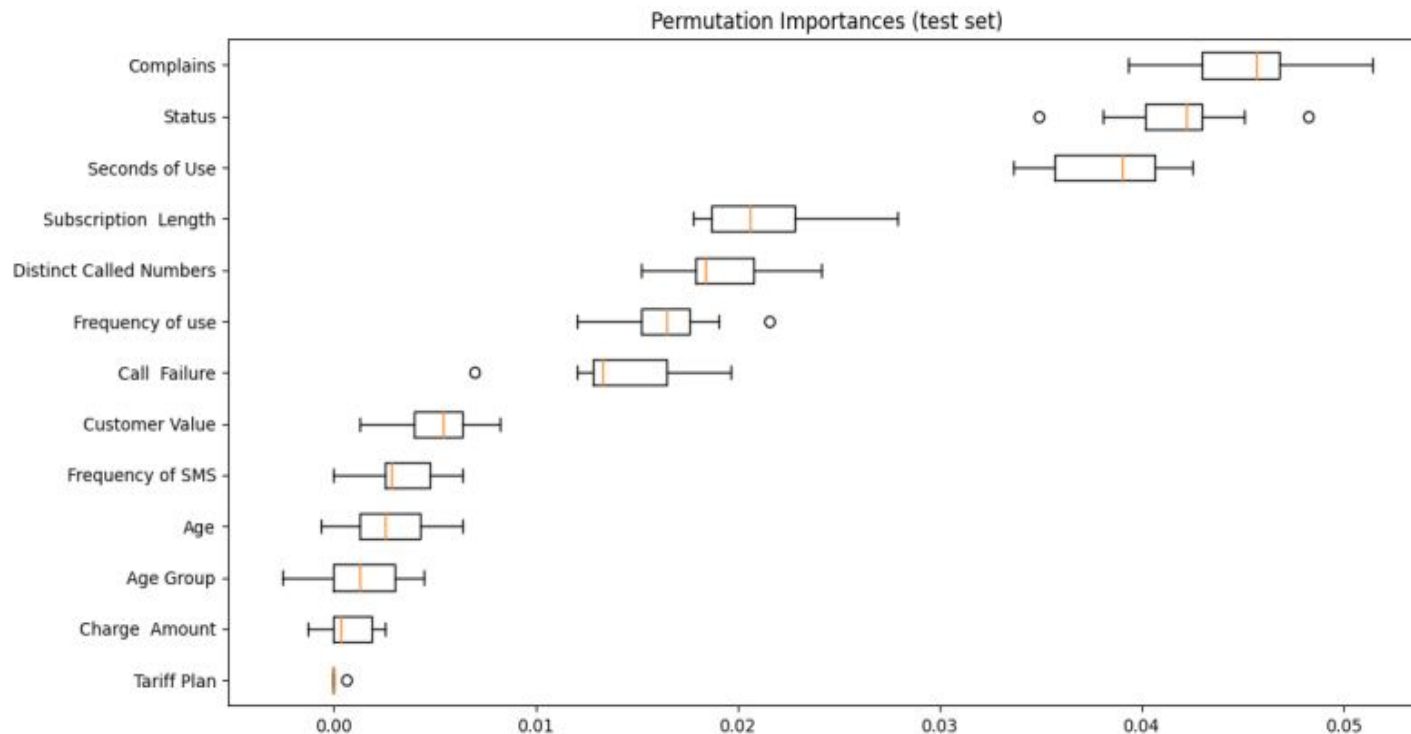
# Optimization for predictability on unseen data



## Model Optimization

- Reduces overfitting
- Continuous enhancement in the model's performance
- Mean cross-validation score: 0.947

# Key Predictors of Customer Churn



## Model Optimization

- Complaints
- Status (Active/Inactive)
- Seconds of use

# Maximizing for business value based on cost matrix

## 1. Assumptions on customer level

Monthly revenue	25,00 €
3 year revenue	900,00 €
Net profit (15%)	135,00 €
Campaign costs (10%)	13,50 €

## 3. Performance evaluation

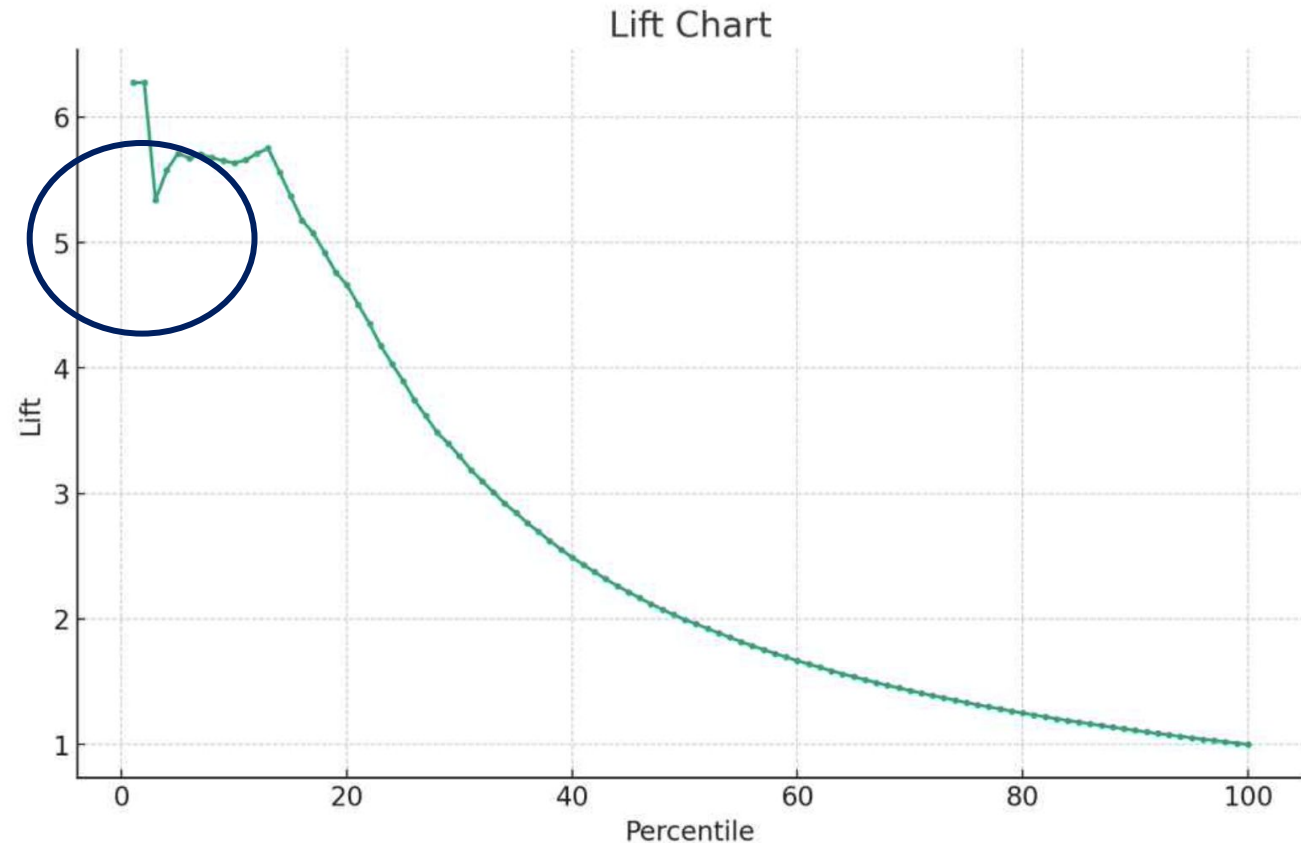
Net profit without model	178.740,00 €
Net profit with model	201.973,50 €
Added value	23.233,50 €
Increase in profit	13,00%

## 2. Assessing the real business impact

Confusion Matrix and Cost analysis			
	Predicted (1)	Predicted (0)	Sum
Actually (1)	194	57	251
Actually (0)	25	1299	1324
Sum	219	1356	1575
Payoff Matrix			
	Predicted (1)	Predicted (0)	
Actually (1)	121,50 €	0	
Actually (0)	- 13,50 €	0	
Payoffs Values			
	23571	0	
	-337,5	0	
Added Value	23.233,50 €		
Added Value / customer / 3 years	14,75 €		

\*assuming 100% effectiveness of the retention strategies

# Lift of 5,66 facilitates effective targeting of probable churners



By targeting the **5% of the churners** with highest probabilities to churn, our model performs **5,66x better** compared to pure chance

# Three proactive strategies to address the prevalent factors of churn



## Communication strategies

- Effectively gather customer feedback with surveys
- Check-in messages and loyalty program
- Personalized benefits



## Continuous improvements

- Analysis of recurring requests
- Addressing high-prio needs
- Service and process optimization



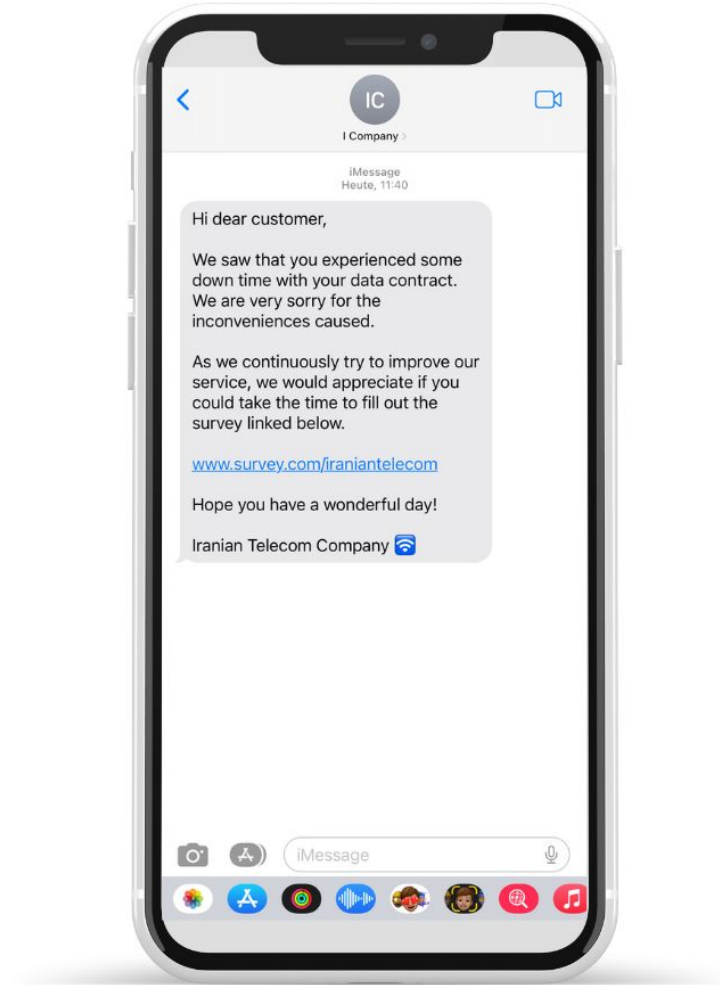
## Subscription terms

- Promote cross and up-selling
- Renegotiation of contracts
- Offer special hardware offers

**Proactively increasing value** for customers (personal cost/benefit analysis) to **prevent churn**



# Potential implementation for service improvement and engagement



# Expected results of the implemented strategies



## Proactive churn prediction

- Being ahead of time  
→ mitigation strategies
- Not jeopardizing margins with last minute discounts



## Financial added value

- Retaining customers more cost-effectively
- Higher CLV
- Higher loyalty → lower CAC



## Service improvement

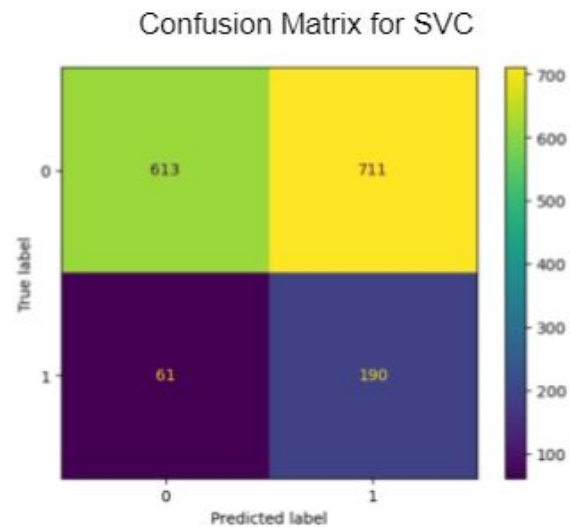
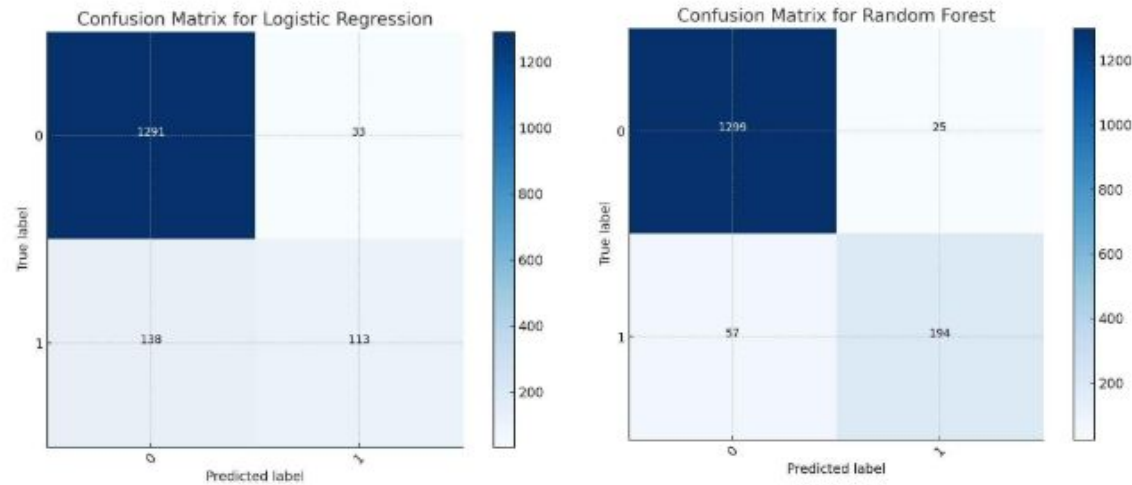
- Higher customer satisfaction
- Long-term competitive advantage

**Great opportunity to improve retention and create value for the company**

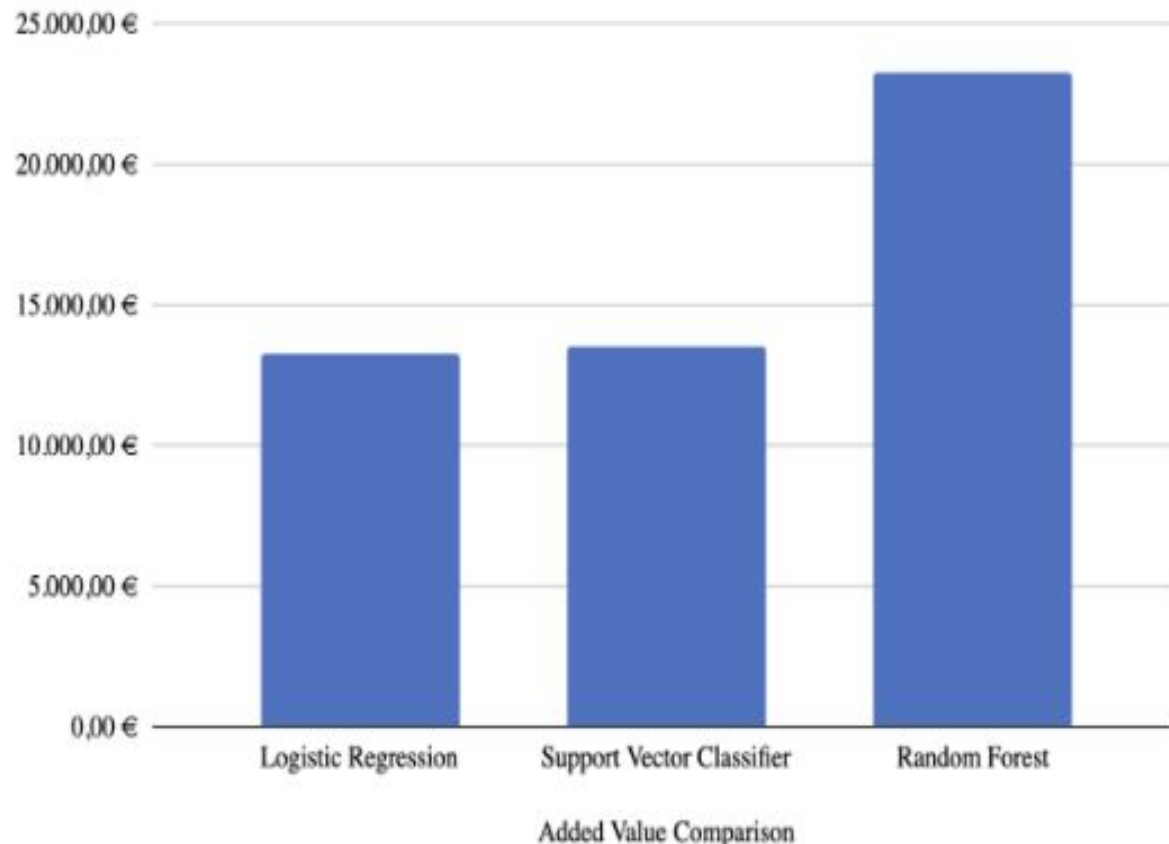
**Thank you**

# Appendices

# Appendix A | Confusion Matrix



## Appendix B | Evaluation of different models on the real business impact



### Reasoning to choose Random Forest as optimized model

1. Best combination of highest lift and highest added value
2. Highest Profitability at optimum threshold with the greatest value of 23.233,50€, exceeding the other two models by approximately 10,000€



# Appendix C | Precision-Recall Curve

- *This curve visualizes the trade-off between precision (the proportion of true positives among all positive predictions) and recall (the proportion of true positives among all actual positives).*
- *By examining the Precision-Recall curve, we can choose a threshold that balances precision and recall in a way that makes sense for our particular problem.*
- *Our model's strong performance is highlighted by its Precision-Recall curve, reinforcing its effectiveness in predicting customer churn.*

