Utilizing Machine Learning to Enhance Risk Assessment at Tamara

Dr. Sobhi Mejjaouli; Noura AlMandeel; Nouf AlSultan and Hind AlBabtain
Department of Industrial Engineering,
College of Engineering,
Alfaisal University, Riyadh 11451, Saudi Arabia;

Abstract

Tamara executives have a constant concern regarding the risks they face. One of them is with regards to defaulting (non-paying) customers, as they must mitigate this risk for their business to survive and grow. By utilizing the petroleum of today (data), they can better mitigate the risk associated with defaulting customers. Deploying Machine learning models can be used to analyze and to determine the risk levels involved in accepting/rejecting customers. This project aims at addressing the problem caused by defaulting customers, which massively have a negative impact on the bottom-line of Tamara.

Keywords
Machine Learning, Risk Assessment and defaulting customers

1. Introduction / Background

A scientific core of an important activity known as risk assessment, which is used for the evaluation of potential sequences (Faustman and Omenn 2008). Risk assessment has become a dominant public policy tool for making choices, based on limited resources (Council 2009). Risk assessment and management was established as a scientific field some 30–40 years Risk assessment and management was established as a scientific field some 30–40 years (Aven 2016).

Risk assessment of financial intermediaries is an area of renewed interest due to the financial crises of the 1980's and 90's (Galindo and Tamayo 2000). After 2007–2008 crisis, it is clear that corporate credit scoring is becoming a key role in credit risk management (Luo, et al. 2017). Consumer credit scoring is often considered a classification task where clients receive either a good or a bad credit status (Kruppa et al. 2013). In this regard (Sun and Vasarhalyi 2021) develop a prediction system to help the credit card issuer model the credit card delinquency risk. Credit card is getting increasingly more famous in budgetary exchanges, simultaneously frauds are likewise expanding (Almuteer et al. 2021).

1.1 About Tamara

Tamara is a Saudi fin-tech startup financial solutions company. It was founded by a team of creative Saudis, every member of which distinguished in their own right, many of whom earned national and global awards to boot. They're the leading Buy Now Pay Later provider in the MENA region, as they provide a Buy Now Pay Later solution for customers to pay with the ability to split their payments into installments with no fees, and no interest. The company operates out of its HQ in Saudi Arabia, and it has offices across the UAE, Germany and Vietnam.

1.2 Vision Statement

Team Tamara works around the clock, constantly developing payment, shopping, and customer experience solutions that make life easier and more fulfilling. Their mission is to empower people to shop through an honest, transparent and inclusive financial solution.

1.3 Artificial Intelligence (AI) and Machine Learning (ML)

AI is a science that empowers computers to mimic human intelligence. Machine Learning is a subfield of AI that enables machines to improve at a given task with experience. Nowadays, AI and ML play a vital role in our daily lives, and examples of applications in our day-to-day lives are: Recommendations of Amazon, Facial Recognition of iPhone, Voice commands of Siri, Routing through Google Maps, Email spam detection, Chatting with

Customer Support (Bots), to even our search engine results on Google.

Problem Statement

"We face issues with our rule-based system that determines whether or not to accept a persons' request"

Problem Background

The risk of defaulting (non-paying) customers is vital to Tamara, as it's at the core of how Tamara operates and generates revenue. To understand the risk levels of different customers, Tamara can utilize the vast amounts of Information they have. Deploying Machine learning models can be used to analyze or to determine the risk levels involved in accepting/rejecting customers. This project aims at addressing the problem caused by defaulting customers, which hugely affects the bottom-line of Tamara.

Data collected

Delinquency Dataset: 10K transactions made by users that paid using Tamara.
The sample data included:

- Installment, Total, Refunded amount
- **Order ID**
- Merchant name
- # of installments
- Product description
- Status

Customer Dataset: A table including the following:

- Customer ID
- **Order ID**
- Customer name
- Customer Gender (Male/Female)
- Employment Status (Employed/Not employed)

	A	B	C	D	E	F	G	H	I	J
1	created_at	Instalment amount	id_order	merchant_name	country_code	description	locale	total_amount	instalments	status
2	2021-01-31 00:24:10	58.4	4f192a74-0484-4728-b5ac-40a4d5a241a3	Johrh	SA	استخدم خدمة تمارا مع ووكومرس	ar	1752000		fully_captured
3	2021-01-31 00:25:27	257.45	a433272b-e341-457c-9418-2fc0265c9167	Namshi	SA	Namshi Order - SA8C8169169	ar_SA	3614500		fully_captured
4	2021-01-31 00:35:36	96.9	27c97813-533d-49a2-a594-89d8a815cdef	Namshi	SA	Namshi Order - SA8C1817715	ar_SA	5158100		fully_captured
5	2021-01-31 00:45:28	256.5	def856c3-23e1-4bd1-858c-9d3257e1709f	Namshi	SA	Namshi Order - SA8C0017913	ar_SA	2565000		fully_captured
6	2021-01-31 01:56:51	262	8cec7a1e-d5fd-43fc-8ea1-ca22e59d6e20	Sharks ksa	SA	طلبك من شاركس | Sharks	ar_SA	2620000		fully_captured
7	2021-01-31 03:21:35	209.35	5703c6c8-cd32-4bb7-b28a-457ab642260b	safqaplus.com	SA	طلبك من صفقة بلس	ar_SA	6290500		fully_captured
8	2021-01-31 05:40:14	68.4	da392d7c-407e-4c75-a93a-8f48dc53b88d	Namshi	SA	Namshi Order - SA8C6045284	ar_SA	3144500		partially_captured
9	2021-01-31 06:05:46	174	cc97c6bc-2620-4bd9-935b-75958b3e2644	Namshi	SA	Namshi Order - SA8C6249541	ar_SA	9244500		fully_captured
10	2021-01-31 06:15:35	237.5	b16d0738-2583-4d87-874b-ea0be66c7f41	Namshi	SA	Namshi Order - SA8C6374636	ar_SA	2375000		fully_captured
11	2021-01-31 06:25:36	515.09	34d8ee91-930e-4d2a-93af-2ee231363268	Namshi	SA	Namshi Order - SA8C6708713	ar_SA	6177000		partially_captured
12	2021-01-31 07:00:14	111.69	591199ea-6c2b-4ff9-b791-612cb651eb2d	Namshi	SA	Namshi Order - SA917698603	en_SA	1478000		partially_captured
13	2021-01-31 07:30:18	328.64	84903983-633e-4b40-8106-7973209cb3db	Namshi	SA	Namshi Order - SA915189005	ar_SA	5608500		partially_captured
14	2021-01-31 07:50:09	358	bce5bfe7-d4d5-4de0-86d8-4b030329c289	under7score.com	SA	طلبك من UnderScore	ar_SA	3580000		fully_captured
15	2021-01-31 08:08:40	446	16db6047-5a84-4393-ac16-c5d43d813605	Toys House SA	SA	طلبك من Toys House	ar_SA	4460000		fully_captured
16	2021-01-31 08:10:27	475	fb40f2ef-919b-4b67-90b7-7d461d3a6a9e	Namshi	SA	Namshi Order - SA918353827	ar_SA	4750000		fully_captured
17	2021-01-31 08:20:22	85	56deddcb-2e6c-48d5-a622-2f2dc2c47925	Namshi	SA	Namshi Order - SA912866616	ar_SA	2660000		fully_captured
18	2021-01-31 08:40:01	85.15	e63ff8b4-bc1e-4d09-9a6f-bc48eb720832	Johrh	SA	استخدم خدمة تمارا مع ووكومرس	ar	6289000		partially_refunded
19	2021-01-31 09:42:14	96	8d4137c3-e85b-4695-b523-6c59216d3a50	Bejeelah	SA	طلبك من Bejeelah	ar_SA	2900000		fully_captured
20	2021-01-31 09:52:43	283	337cdc4b-f691-4617-bc46-8d55bc55a19e	Dar Lena	SA	طلبك من دار لينا	ar_SA	8490000		fully_captured
21	2021-01-31 10:15:16	184.85	cbe9227f-1ded-46cb-8a79-5de5b665ece0	Namshi	SA	Namshi Order - SA917709548	ar_SA	1848500		fully_captured
22	2021-01-31 10:40:43	216.6	1635b15a-9987-45e6-848f-8c56298028c8	Namshi	SA	Namshi Order - SA913092896	ar_SA	2166000		fully_captured
23	2021-01-31 10:40:45	721	69c2b6a2-14f8-4a65-8afb-6b49783b3c15	Namshi	SA	Namshi Order - SA910807199	en_SA	8740000		fully_captured
24	2021-01-31 10:45:16	356.2	81b5c8c7-3f43-4052-8d13-adc00719a8b1	Namshi	SA	Namshi Order - SA913887383	ar_SA	5842000		fully_captured
25	2021-01-31 10:48:51	143	98028c54-5db8-4681-8b0e-ac7d5bd0dce6	Sharks ksa	SA	طلبك من شاركس | Sharks	ar_SA	1430000		fully_captured

Figure 1. Collected data

Methodology

The proposed technique is using a common machine learning model "Naïve Bayes Classifier".
The idea is to study the characteristics of different customers, then select the most vital characteristics/features that that has an effect on whether or not the customer defaults (fails to pay on time). The most vital features were:

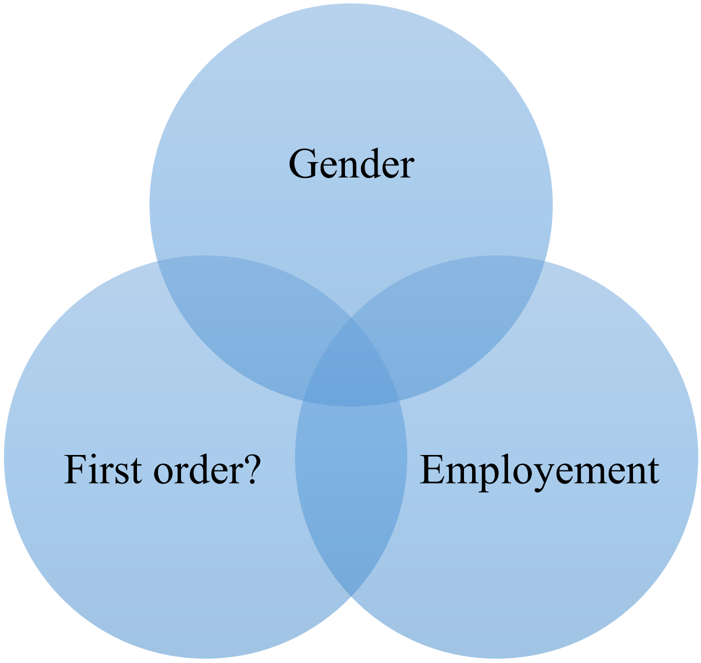


Figure 2. characteristics of customers

Then, we randomly split our dataset utilizing the Train/Test method to helps us to measure the accuracy of our model later, it is common to use 80% of the data for training, and the rest 20% for testing, since these unknown data will help us ensure that our model yields great results. Subsequently, we train the dataset to further understand the key variables/features. The equation for the Naïve Bayes is:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$P(c|x)$_*Likelihood of the evidence given that the hypothesis is true*
$P(x|c)$_*Posterior probability of the hypothesis given that the evidence is true*
$P(x)$_*Prior probability of the hypothesis*
$P(c)$_*Prior probability that the evidence is true*

We then test the model on the remaining 20% of the data and check its performance.
Afterwards, we build a rule-based system to accept/reject customers. Ultimately, we perform calculations of the money that could've been saved if this system had been implemented.

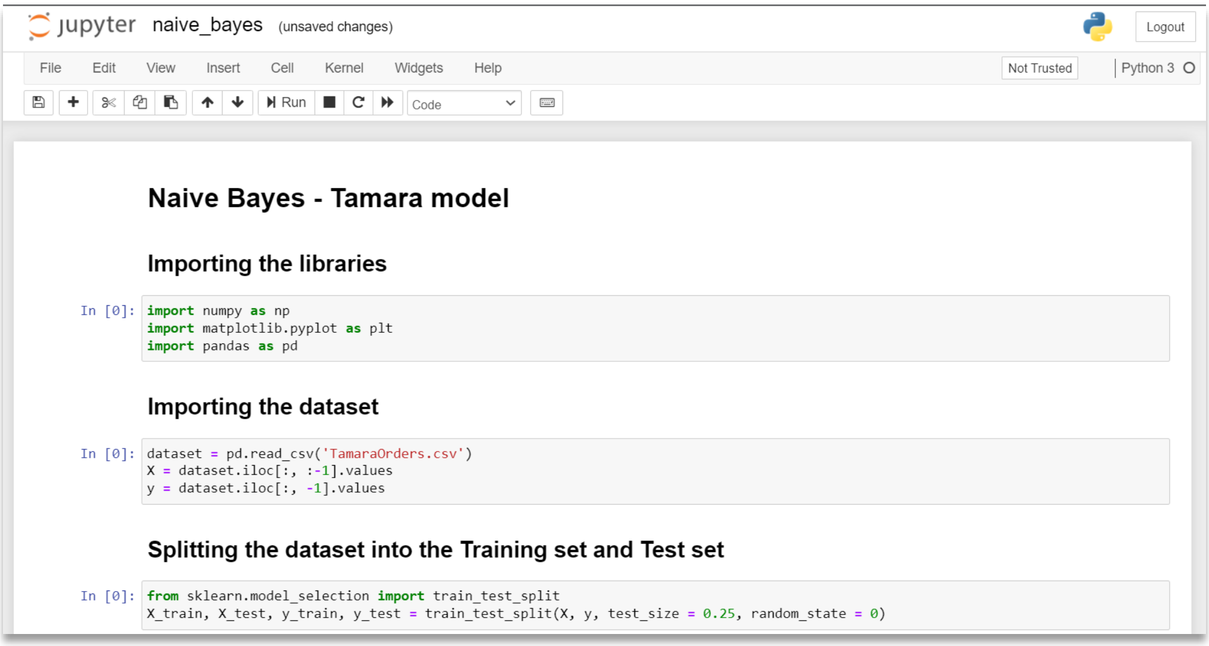


Figure 3. Screenshot of the Python code used to run the model

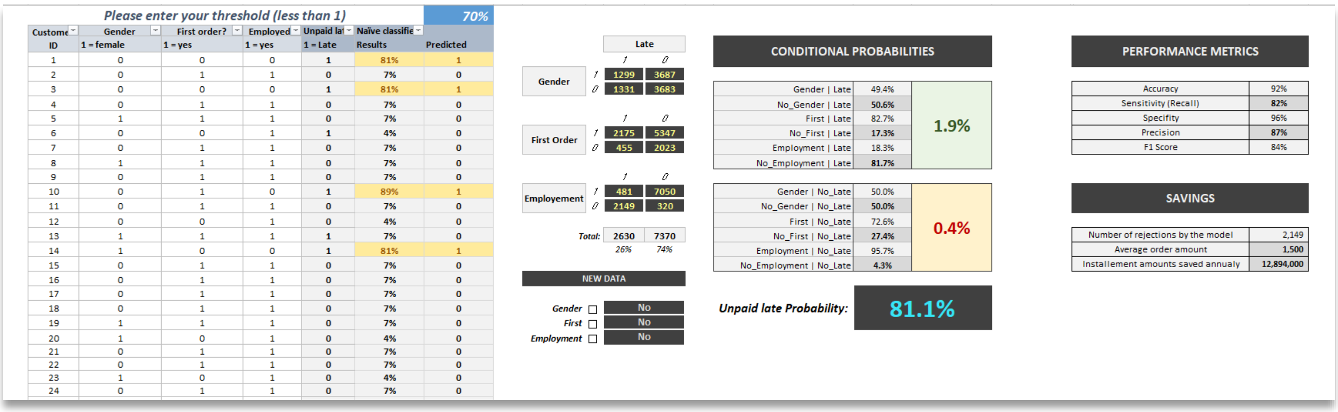


Figure 4. Screenshot of the Excel template developed to run the model

The steps to use the excel template are explained in this section.
The user prompted to enter their threshold (less than 1) – which is a percentage of the tolerance they can have before concluding whether a customer is likely to default or not. If this is 60% then there is a big room for the model to accept customers who eventually turn out to be defaulting customers. At the same time, making the percentage near 100% almost makes it very difficult to accept new customers. Therefore, a threshold percentage of 70% here was recommended and deemed acceptable. Afterwards, the data for the three features are decoded into 1 and 0s only. The following decoding was performed:

Table 1. decoding

Gender	First order?	Employed?
1 = female	1 = yes	1 = yes
0 = male	0 = no	0 = no

Then after inserting all the values for the sample data of 10,000 rows which is equivalent to the orders performed in a quarter of the year. The column of Unpaid Late indicates the true value for this order and whether or not this customer has paid or not. The column of Naïve classifier returns a probability based on the model we designed. This probability / percentage is compared to the threshold the user enters as mentioned above. Thus, the last column of this data set is the result of the prediction.

Table 2. excel template

Please enter your threshold (less than 1)						70%
Custome ID	Gender 1 = female	First order? 1 = yes	Employed 1 = yes	Unpaid la 1 = Late	Naïve classifie Results	Predicted
1	0	0	0	1	81%	1
2	0	1	1	0	7%	0
3	0	0	0	1	81%	1
4	0	1	1	0	7%	0
5	1	1	1	0	7%	0
6	0	0	1	1	4%	0
7	0	1	1	0	7%	0
8	1	1	1	0	7%	0
9	0	1	1	0	7%	0
10	0	1	0	1	89%	1
11	0	1	1	0	7%	0
12	0	0	1	0	4%	0
13	1	1	1	1	7%	0
14	1	0	0	1	81%	1
15	0	1	1	0	7%	0
16	0	1	1	0	7%	0
17	0	1	1	0	7%	0
18	0	1	1	0	7%	0
19	1	1	1	0	7%	0
20	1	0	1	0	4%	0

The next part of the model involves all the calculations for the probabilities.

Table 3. calculations

		Late 1	Late 0
Gender	1	1299	3687
	0	1331	3683
First Order	1	2175	5347
	0	455	2023
Employement	1	481	7050
	0	2149	320
	Total:	2630	7370
		26%	74%

NEW DATA	
Gender ☐	No
First ☐	No
Employment ☐	No

The columns and rows in here represent the intersections of the different variables vs the main outcome which is if the customer is late or not. This is used in calculating the conditional probabilities as shown here. And for each different combination of variables, this is being computed. Here is the value for when all of the variables = 0.

Table 4.Conditional probabilities

CONDITIONAL PROBABILITIES		
Gender | Late	49.4%	1.9%
No_Gender | Late	**50.6%**	
First | Late	82.7%	
No_First | Late	**17.3%**	
Employment | Late	18.3%	
No_Employment | Late	**81.7%**	
Gender | No_Late	50.0%	0.4%
No_Gender | No_Late	**50.0%**	
First | No_Late	72.6%	
No_First | No_Late	**27.4%**	
Employment | No_Late	95.7%	
No_Employment | No_Late	**4.3%**	
Unpaid late Probability:	81.1%	

And here are the results of all the different combinations.

Table 5. Results of combinations

	Gend	First orde	Employe	Probability
000	0	0	0	81.08%
010	0	1	0	88.57%
001	0	0	1	4.17%
011	0	1	1	7.30%
100	1	0	0	80.69%
110	1	1	0	88.31%
101	1	0	1	4.07%
111	1	1	1	7.13%

Furthermore, to ensure the reliability of our model, we calculated the most common performance metrics. Each one is utilized differently. They're explained here in the next section. Performance Metrics (Confusion Matrix):

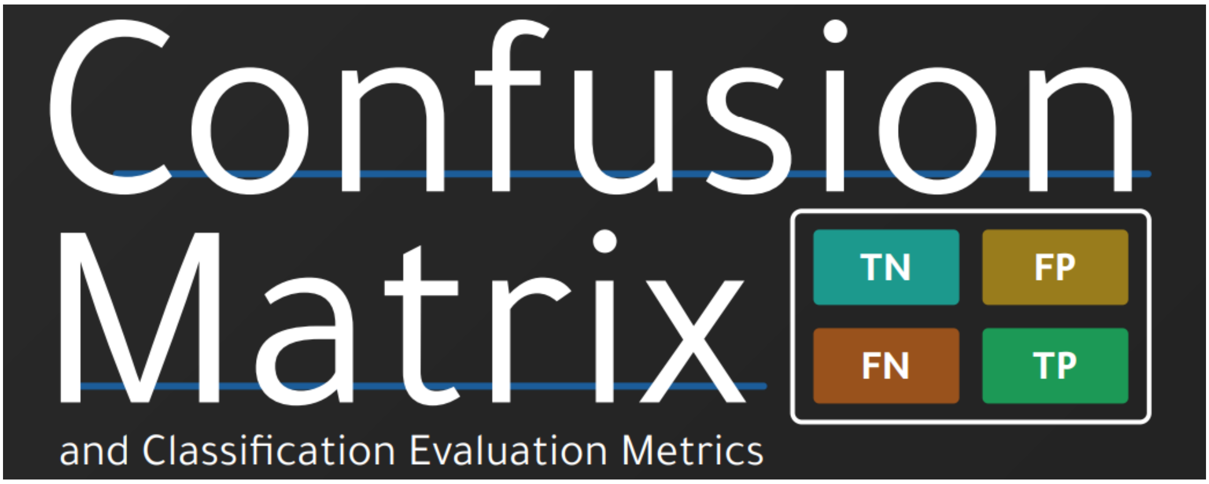


Figure 5. Performance Metrics (Confusion Matrix):

Trust is a must when a decision-maker's judgment is critical. To give such trust, we summarize all possible decision outcomes into four categories: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) to serve an outlook of how confused their judgments are, namely, the confusion matrix. From the confusion matrix, we calculate different metrics to measure the quality of the outcomes. These measures influence how much trust we should give to the decision-maker (classifier) in particular use cases. This document will discuss the most common classification evaluation metrics, their focuses, and their limitations in a straightforward and informative manner.

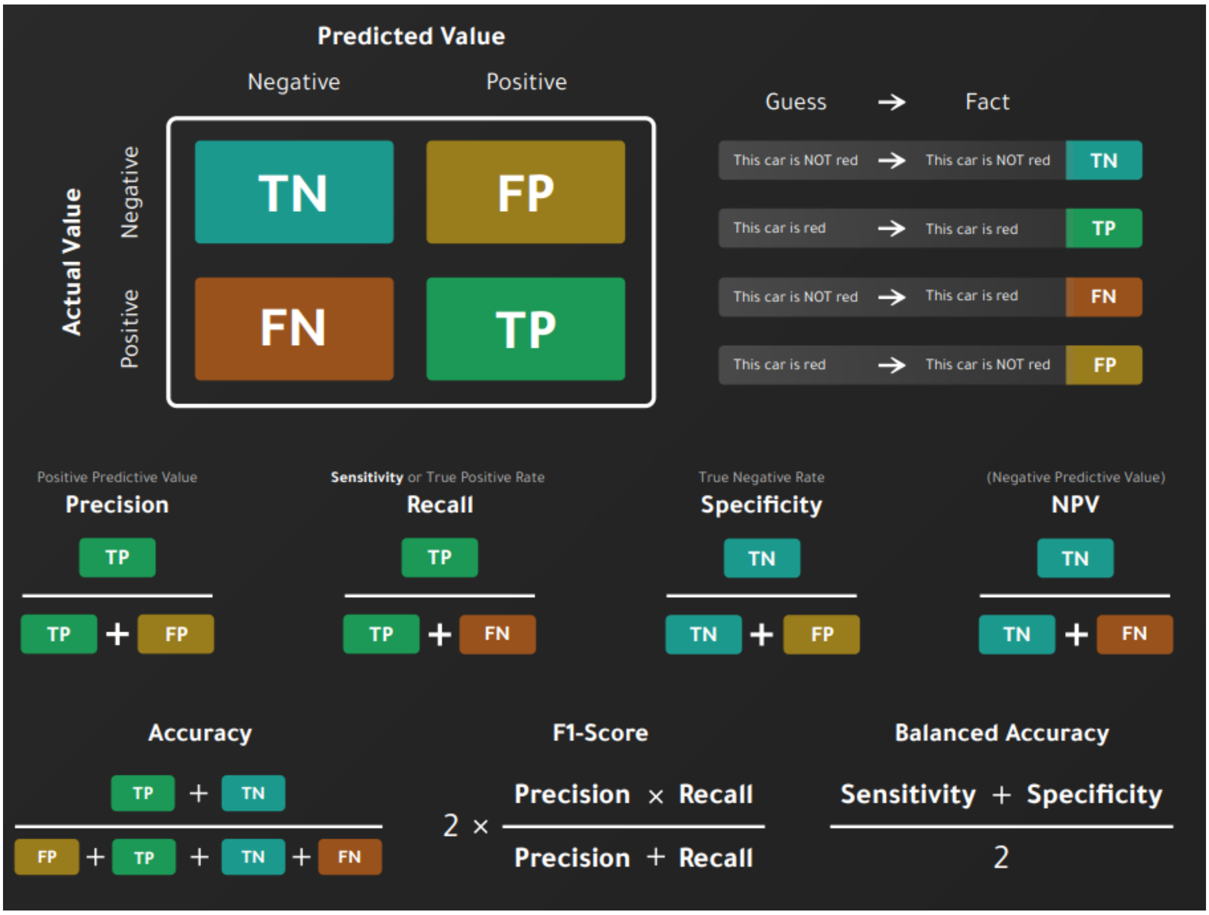


Figure 6. Possible decision outcomes

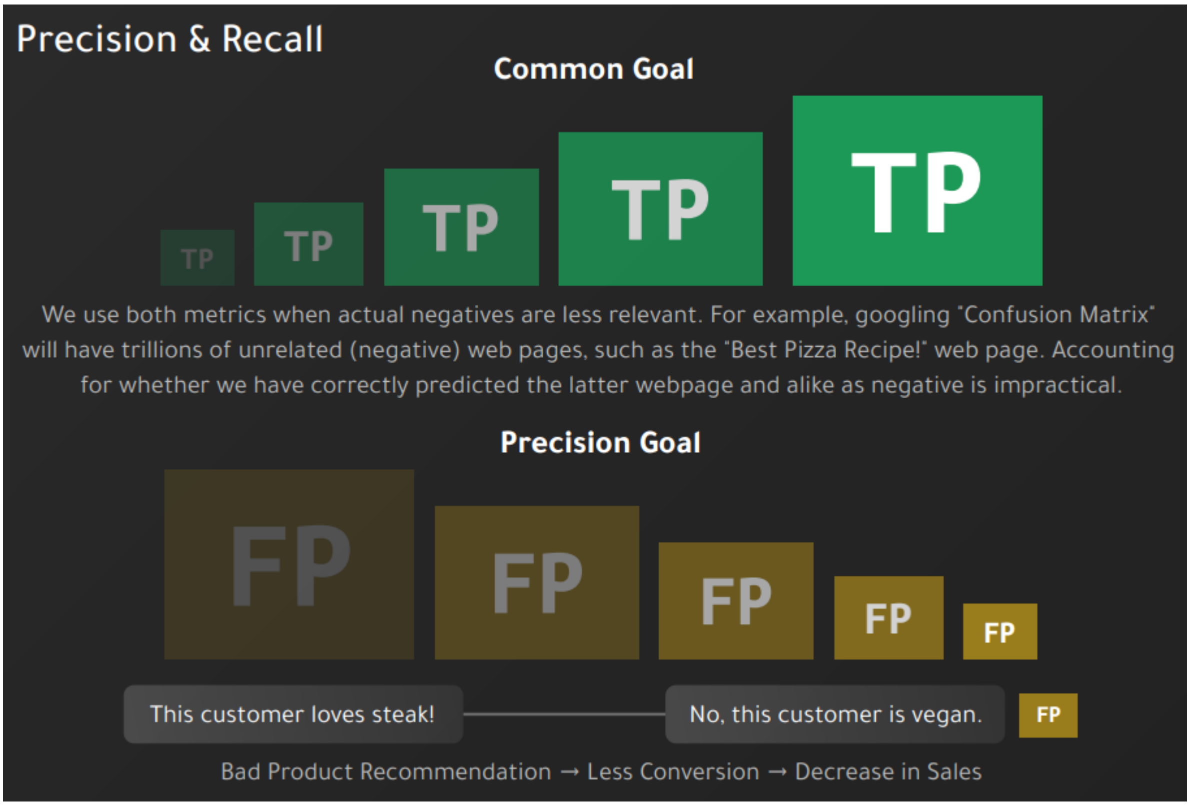


Figure 7. Precision and recall

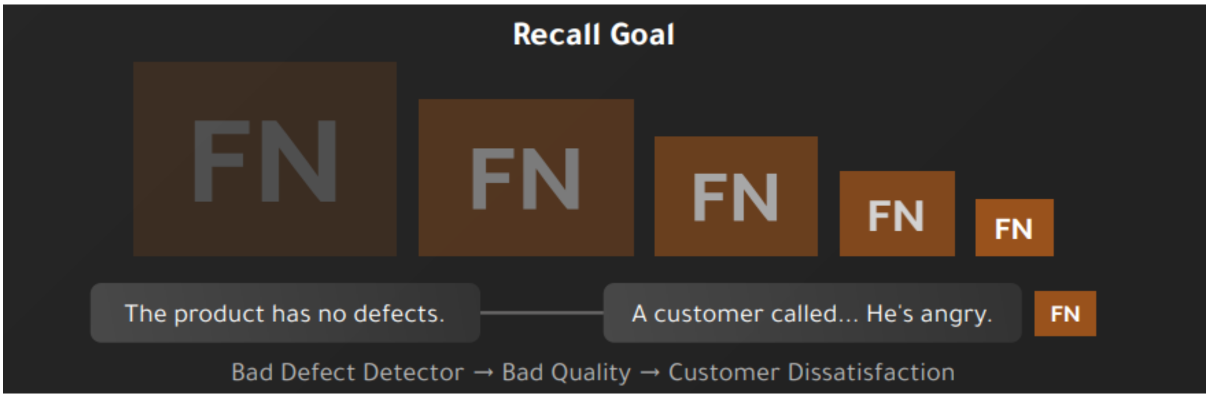


Figure 8: Recall goal

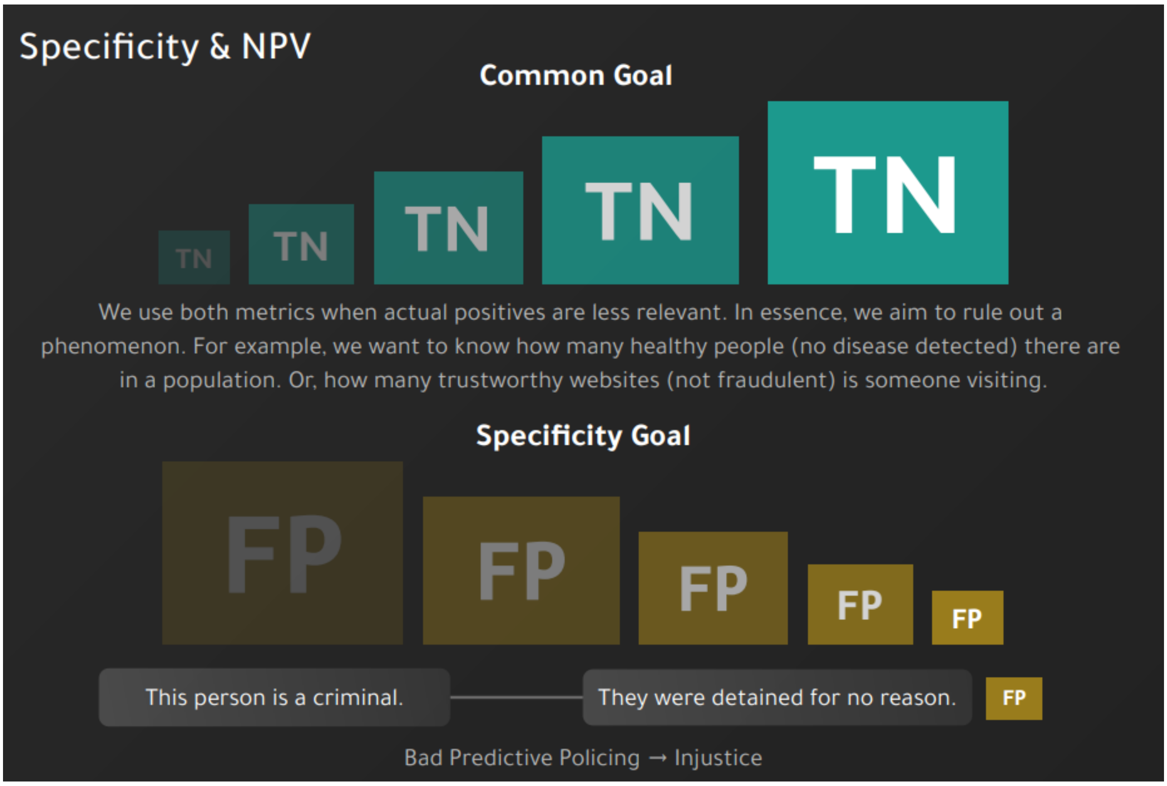
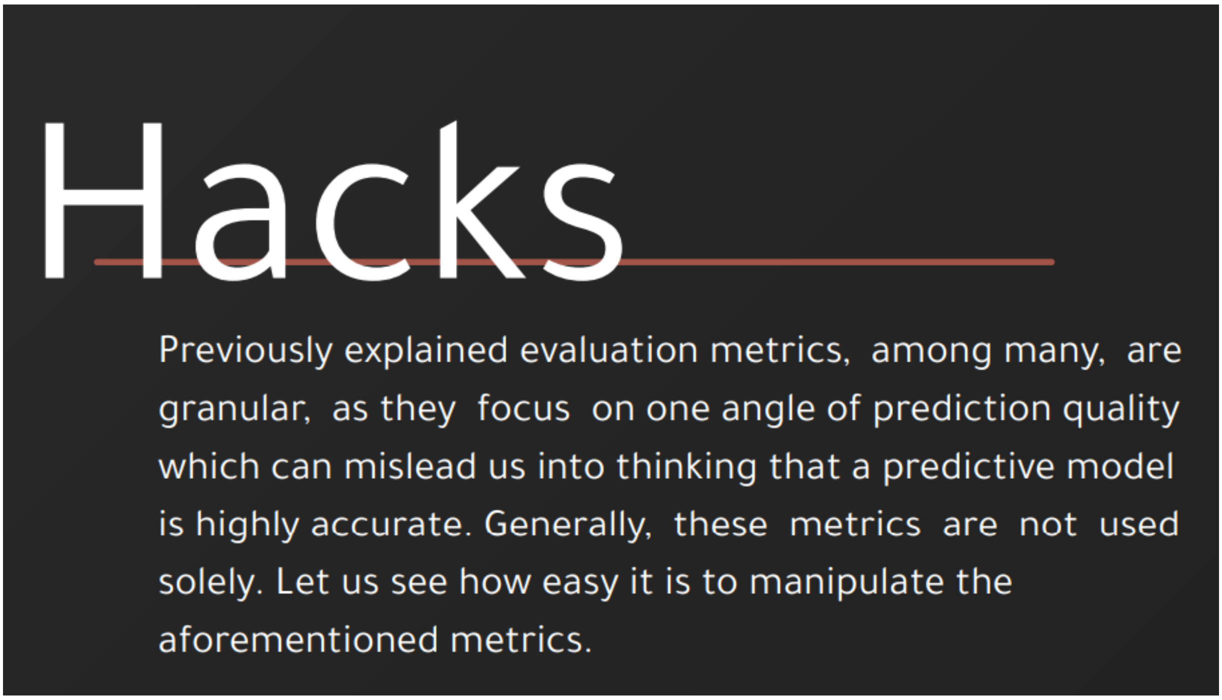


Figure 9: Specificity and NPV


Hacks
Previously explained evaluation metrics, among many, are granular, as they focus on one angle of prediction quality which can mislead us into thinking that a predictive model is highly accurate. Generally, these metrics are not used solely. Let us see how easy it is to manipulate the aforementioned metrics.

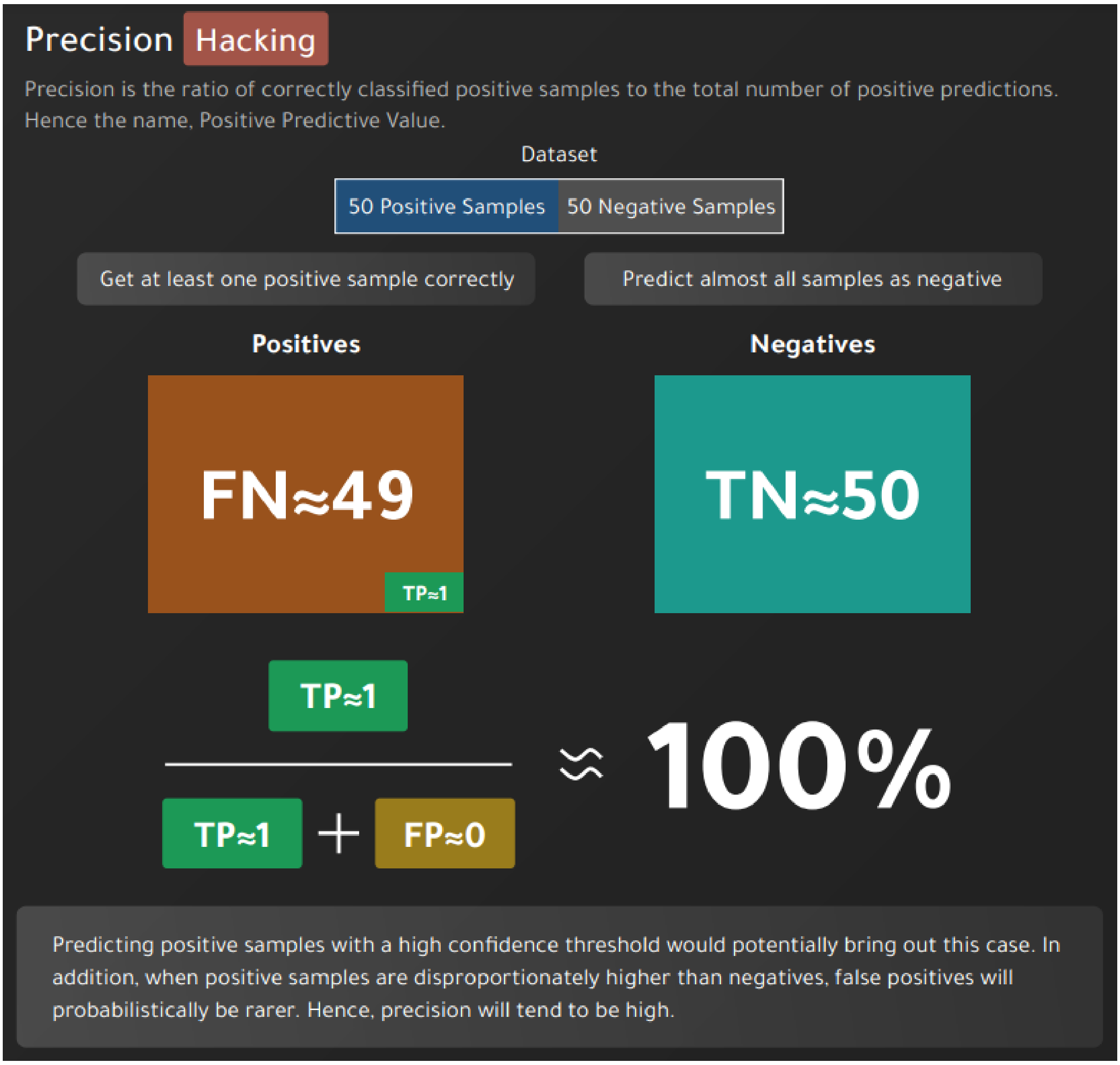


Figure 10, Precision hacking

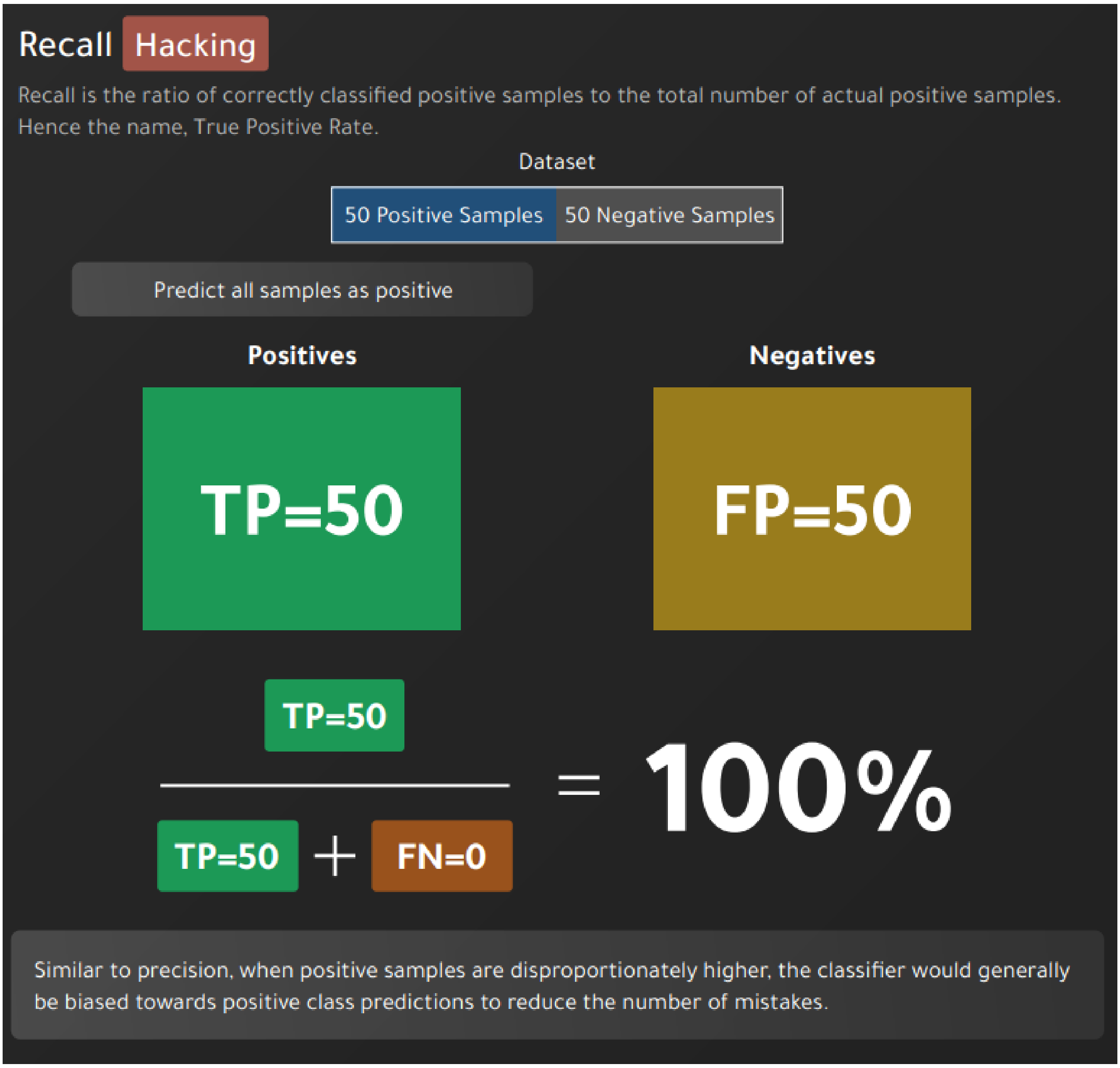


Figure 11. Recall hacking

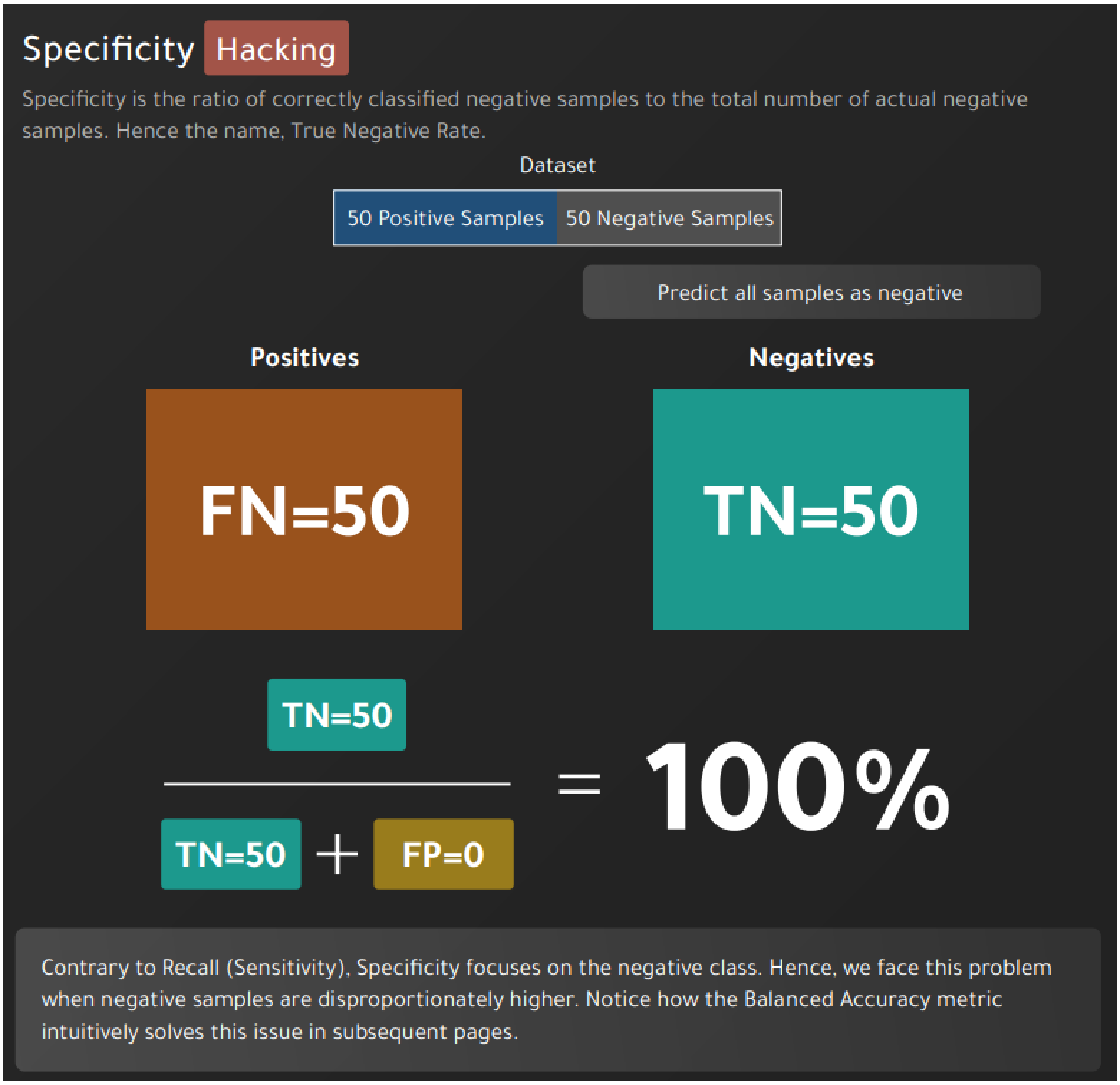


Figure 12. Specificity hacking

Comprehensive Metrics

As we have seen above, some metrics can misinform us about the actual performance of a classifier. However, there are other metrics that include more information about the performance. Nevertheless, all metrics can be "hacked" in one way or another. Hence, we commonly report multiple metrics to observe multiple viewpoints of the model's performance.

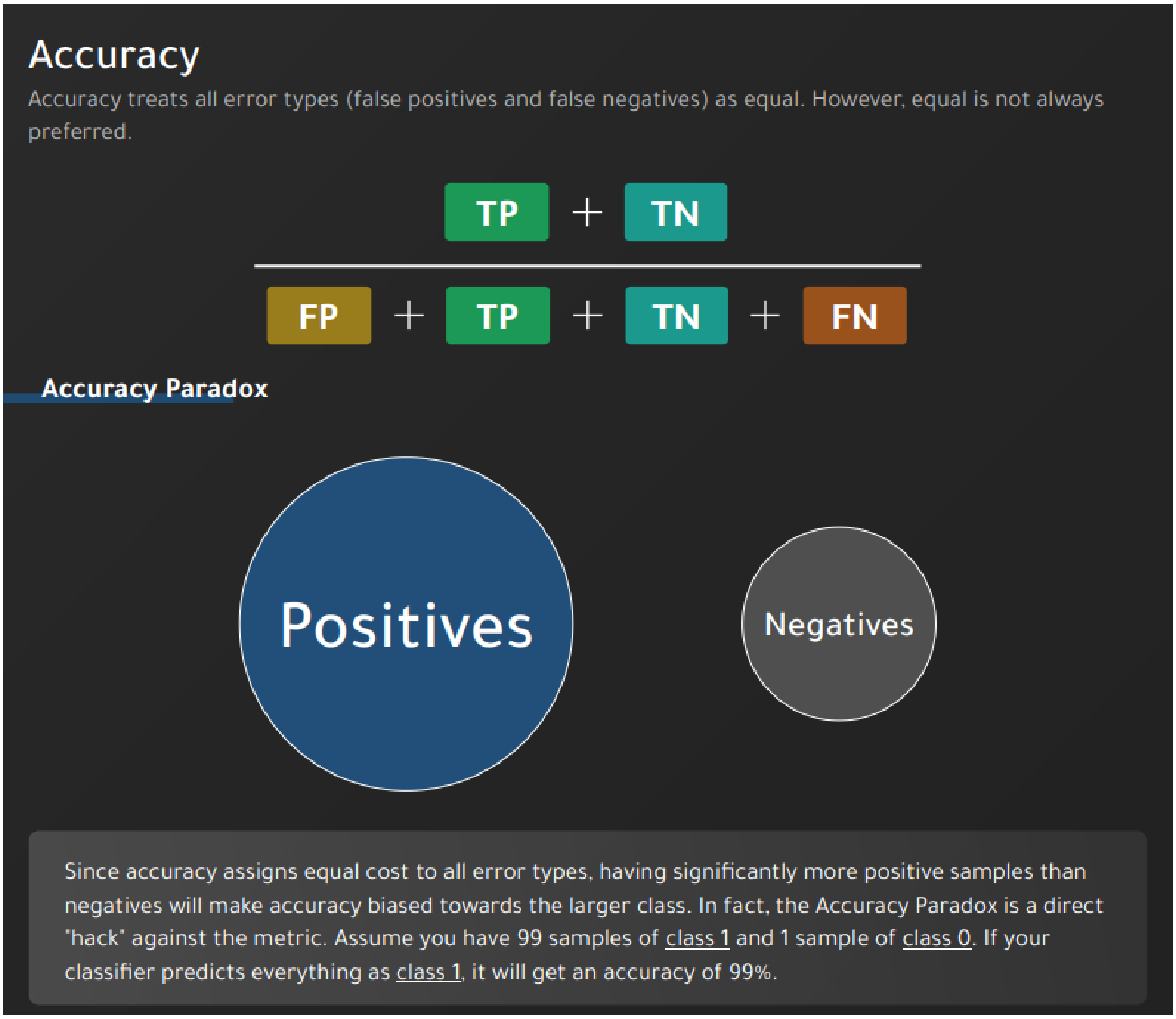


Figure 13. Accuracy

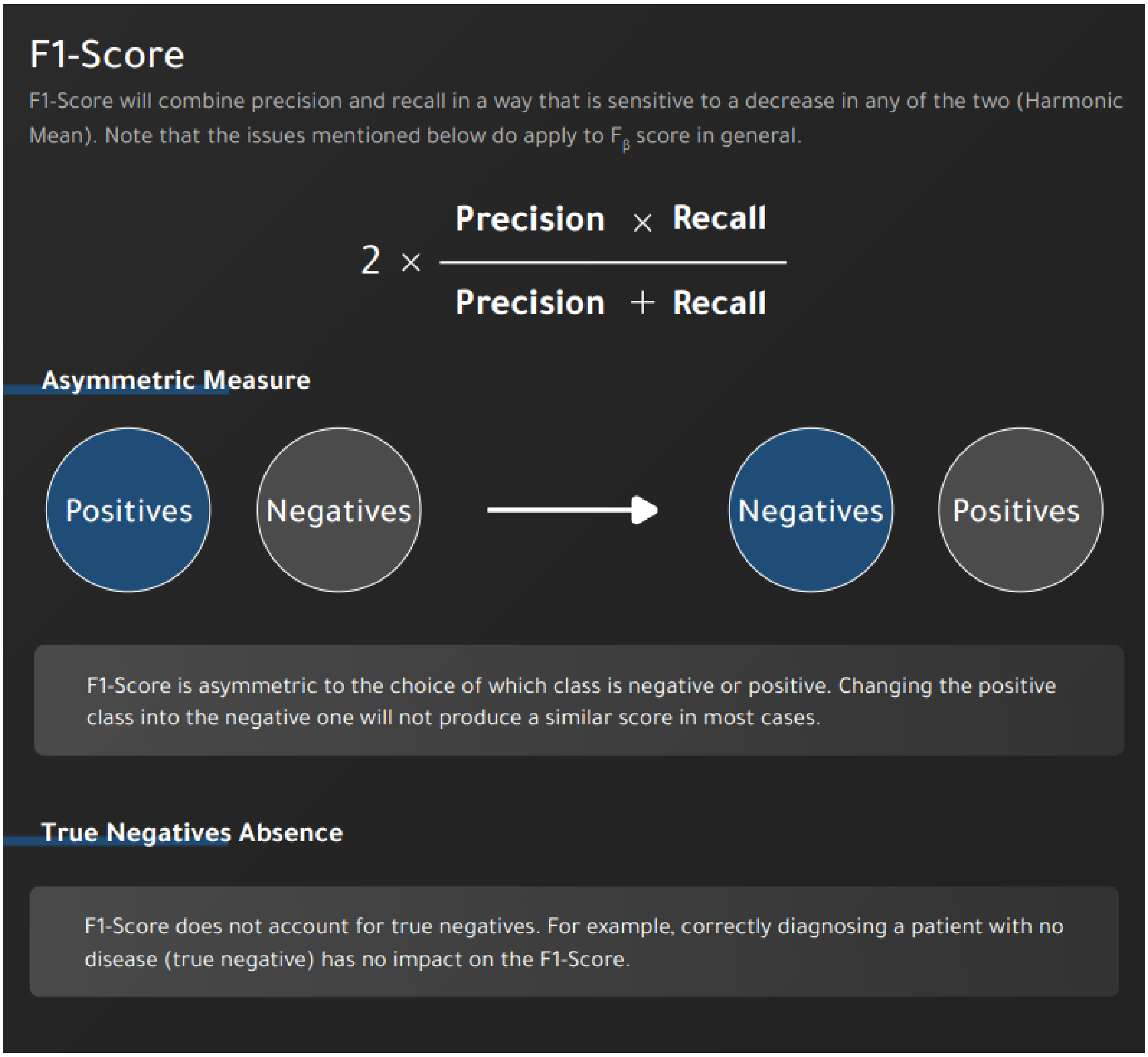


Figure 14. F1-Score

And here are the results we gathered. All the indications are that the model is reliable, and the results generated are actually close to the real situation which gives us assurance and confidence in our work, and that we could implement this going forward.

Table 6: Performance metrics results

PERFORMANCE METRICS	
Accuracy	92%
Sensitivity (Recall)	**82%**
Specifity	96%
Precision	**87%**
F1 Score	84%

		Actual 1	Actual 0
Predicted	1	2149	320
	0	481	7050

SAVINGS	
Number of rejections by the model	2,149
Average order amount	**1,500**
Installement amounts saved annualy	**12,894,000**

Finally, the savings predicted due to this model is calculated. We calculate the number of rejections made by the model. (i.e., when the prediction is the same as the actual and both =1 which is that the customer is late). This means that since all the 10,000 were accepted before, we should have rejected 2,149 of them based on the model. Also, that means that of, course the model didn't predict everything right, but at least the margin of error has been reduced drastically.
Following this number, and based on the statistical analysis of the data, we found that the average order original amount is 1,500 SR. Multiplying these numbers together shows the money that could have been saved in a quarter of the year. Therefore, we multiply again by 4 to find out the expected savings annually.

Conducted Meetings

Kick off meeting 27 October Wednesday:
- Visited the Tamara Main office in Riyadh and attended an introductory meeting about Tamara's main responsibilities and roles.
- As a team, we gave them an overview about our capstone project and aligned on the main objective.

1st meeting 2nd of November Monday:
Tamara's CEO recommended to work on a competitive analysis report and conduct a full market research. The report should include a comparison between Tamara and the other players in the market with the same Buy Now Pay Later concept and find ways to increase the competitive advantage.

2nd meeting 9th of November Monday:
In the second meeting we defined the project scope as a team and proposed a methodology to follow. In summary the methodology included: current state assessment, research and analysis, implementation, and results and outcomes.

3rd meeting 16th of November Monday:
In this meeting we proposed to change the project scope into a scope which is related to what we are studying.

4th meeting 6th of December Monday:
Tamara's CEO shared a sample of the delinquency dataset that contains more than 10K transactions.

Results & Conclusion

Earlier, we have successfully conducted an Exploratory Data Analysis for a sample of the data and gathered some helpful insights that drew our attention.

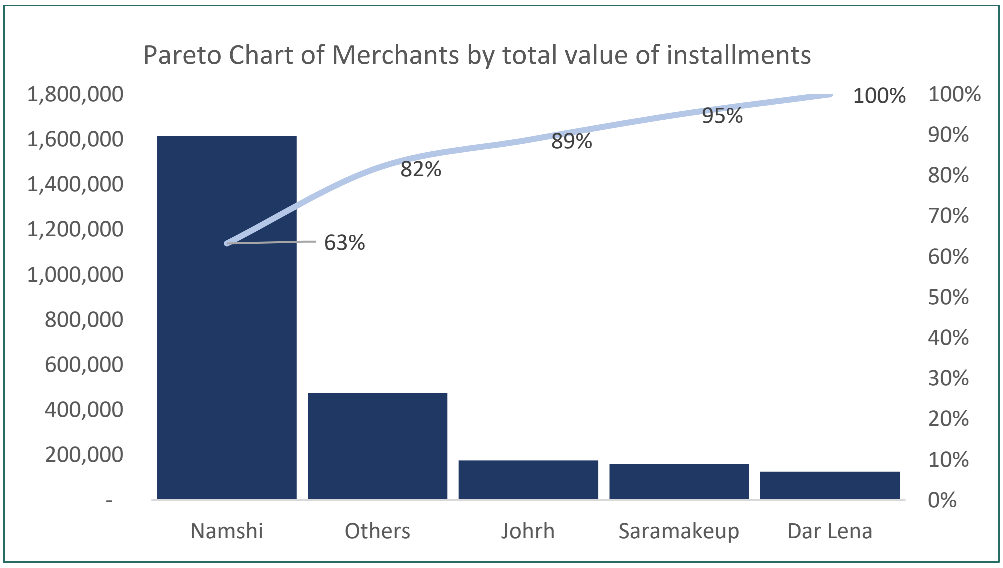


Figure 15: Pareto Chart of Merchants by total value of installments

Based on the Pareto (80-20) rule, only 4 merchants' control over than 80% of Tamara's transactions, with Namshi alone consisting of 63% of it.

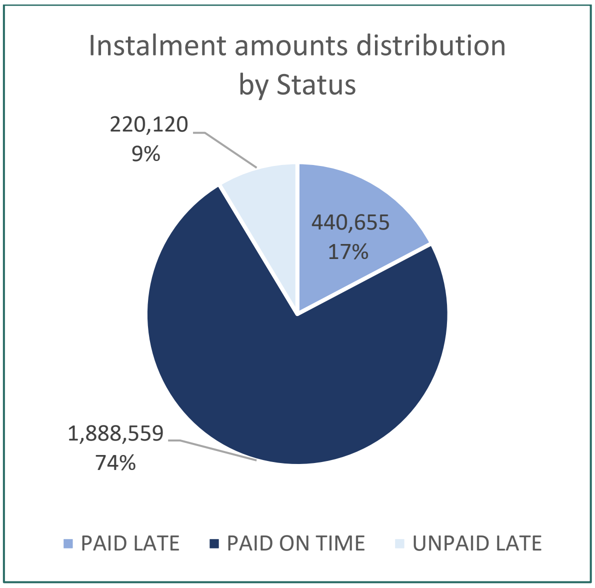


Figure 16. Instalment amounts distribution

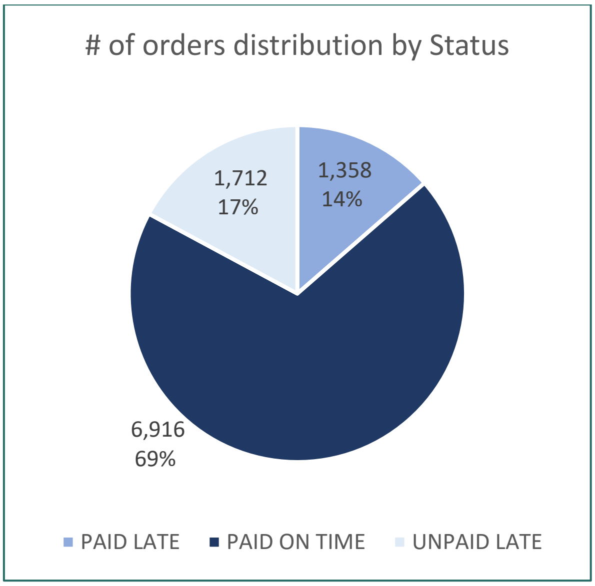


Figure 17. # of orders distribution by Status

- 17% of installments amounts were paid late by Tamara's customers, while 9% are still unpaid.
- 14% of orders were paid late, while 17% are still unpaid.

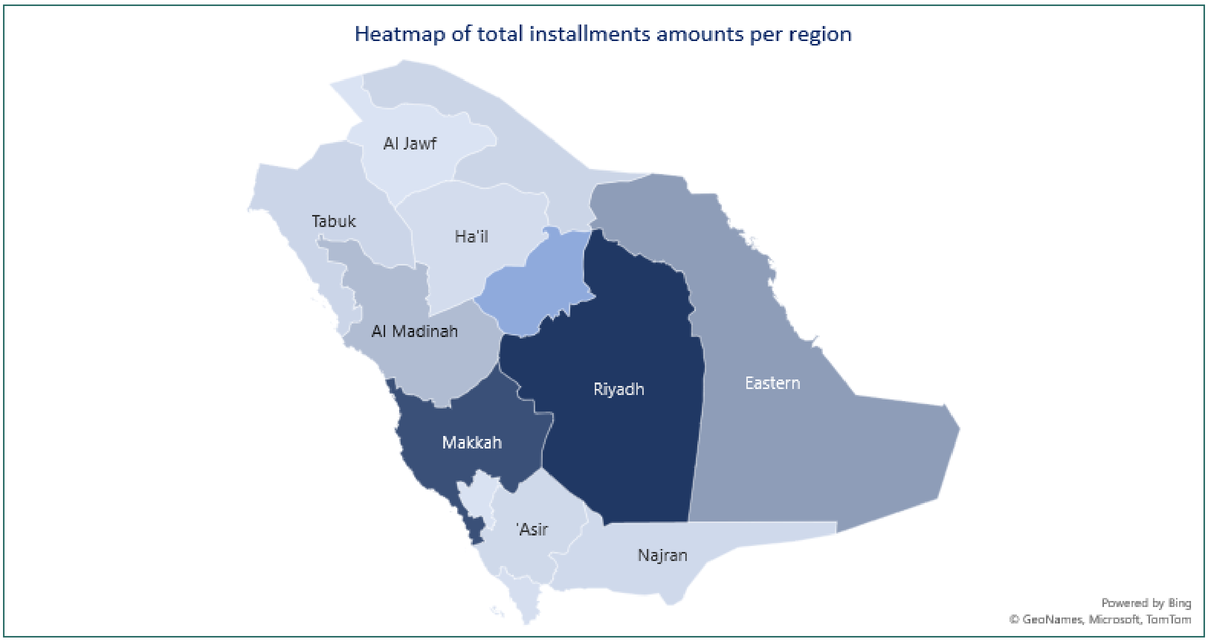


Figure 18. Heatmap of total installments amounts per region.

The concentration of installments amounts by region is based on Riyadh's region, followed by Western Province and then Eastern Province.

Table 7: Savings

SAVINGS	
Number of rejections by the model	2,149
Average order amount	**1,500**
Installement amounts saved annualy	**12,894,000**

At the end of our project, the key results of the savings are shown here. The model has great cost-saving impact. The model has managed to save 82% of the over-due amounts by the users, The amount is equivalent to around 13 million SR. The method utilized has shown a massive improvement in the decision-making process and deploying it shall provide great impact and saves substantial amounts of money.

References

Almuteer, Arjwan H, et al., Detecting Credit Card Fraud using Machine Learning. International Journal of Interactive Mobile Technologies, vol. 15, no.24, 2021.
Aven, Terje.,Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research 253(1):1-13, 2016.
Council, National Research.Science and decisions: advancing risk assessment, 2009.
Faustman, Elaine M, and Gilbert S Omenn.,Risk assessment. Casarett and Doull's toxicology: The basic science of poisons:107-128, 2008.
Galindo, Jorge, and Pablo Tamayo., Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. Computational economics vol. 15, no.1, pp:107-143, 2000.
Kruppa, Jochen, et al., Consumer credit risk: Individual probability estimates using machine learning. Expert Systems with Applications 40(13):5125-5131, 2013.
Luo, Cuicui, Desheng Wu, and Dexiang Wu., A deep learning approach for credit scoring using credit default swaps. Engineering Applications of Artificial Intelligence 65:465-470, 2017.
Sun, Ting, and Miklos A Vasarhalyi, Predicting credit card delinquencies: An application of deep neural networks. *In* Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning. Pp. 4349-4381: World Scientific, 2021.