

Incident Impact Prediction

PoC





Goal statement and problems description

GOAL:

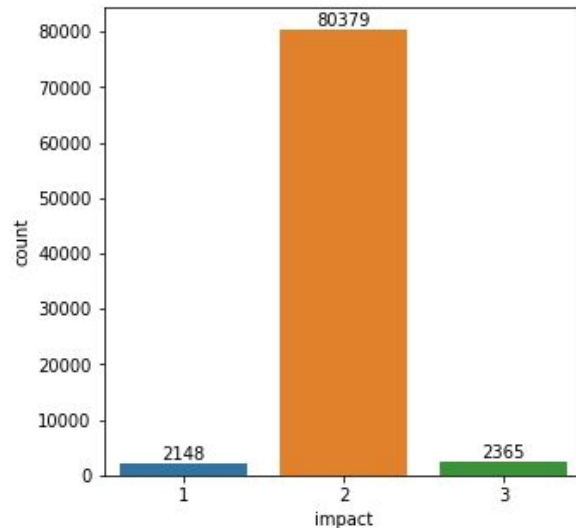
-Predict Impact value for incidents raised by customers

Problems:

-There are a lot of missing data, some columns are almost only missing values (change_request, problem_ID)

-Unbalanced data set - most of Impact values (~95%) are Medium Impact (2). We have only 2.5% of rows with High Impact value (1) and 2.8% of Low Impact values (3)

-There are many features, what features are important for Impact prediction?





Approach and Project Benefits

Approach:

- Analyze and balance the data set, replace and fill missing values, select only important features for impact prediction.
- Use different Machine Learning Models to predict Impact Value.
- This is Multiclass classification Problem.
- Hyperparameter Optimization for best model

Project Benefits:

- Faster High Impact Incidents detection, average incident solution time will be decreased.
- Less manual tasks with labeling for support, they can be more focused on incidents solutions.
- Higher User Satisfaction due to faster solutions and more responsiveness from support.



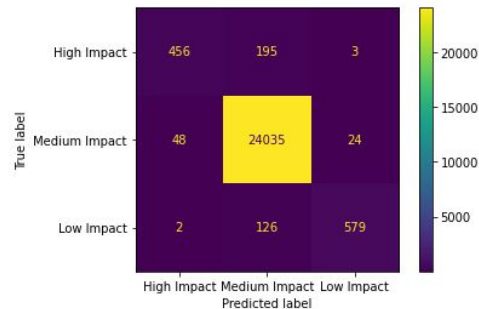
Results

This PoC evaluated four different Machine Learning Models for Multiclass classification:

- Decision Tree Classifier
- Gradient Boosting Classifier
- K Nearest Neighbor Classifier
- Random Forest Classifier

Out of those four models best results were achieved for **Random Forest Classifier**

F1_Score: 88.67%, Accuracy: 98.44%



	Model	data	f1	accuracy	precision	recall
0	Random Forest Optimized	Base_reduced_features	88.67	98.44	94.78	83.77
1	Random Forest	Base_reduced_features	84.77	98.01	95.16	77.65
2	Decision Tree	Base_reduced_features	83.40	97.43	82.66	84.24
3	Decision Tree	Base	83.10	97.35	81.68	84.67
4	Random Forest	Balanced_smote_reduced_feauters	82.14	97.20	82.42	82.86
5	Random Forest	Base	81.94	97.74	95.20	73.79
6	Random Forest	Balanced_smote	80.45	97.21	84.84	77.43
7	Random Forest	Balanced_reduced_feauters	76.78	95.75	72.05	84.47
8	Random Forest	Balanced	75.80	95.78	72.42	81.24
9	Decision Tree	Balanced_smote_reduced_feauters	72.73	94.80	66.86	81.90
10	Decision Tree	Balanced_smote	68.23	93.77	62.85	77.12
11	Decision Tree	Balanced_reduced_feauters	63.94	91.41	56.73	80.60
12	Decision Tree	Balanced	59.46	89.48	52.75	77.67
13	KNeighbors	Base_reduced_features	56.54	94.92	67.80	51.09
14	KNeighbors	Base	56.22	94.90	67.29	50.84
15	KNeighbors	Balanced_smote	55.28	86.58	49.27	77.27
16	KNeighbors	Balanced_smote_reduced_feauters	53.38	85.33	47.80	76.35
17	Gradien boosting	Balanced_smote	49.91	83.08	45.49	71.54
18	Gradien boosting	Base	48.31	95.25	79.21	43.07
19	Gradien boosting	Balanced_smote_reduced_feauters	47.79	79.51	44.29	73.36
20	Gradien boosting	Balanced	47.43	79.35	44.02	73.12
21	KNeighbors	Balanced_reduced_feauters	47.26	78.81	43.89	75.76
22	Gradien boosting	Base_reduced_features	46.99	95.14	79.82	42.12
23	KNeighbors	Balanced	46.94	78.27	43.77	75.32
24	Gradien boosting	Balanced_reduced_feauters	45.34	76.04	42.96	73.92



Next Steps and potential solutions

- Research and try to better balance the data set.
- Try more models, acquire more data and try to create solution with Neural Networks
- Optimize Hyperparameters for more ML models.

Other problems that can be potentially solved with ML:

- Automatically assign support group and support in charge for incidents.
- Auto confirmation check
- Gather additional columns with Incident description and use Natural Language Processing Models to automatically suggest solution based on knowledge base document.

Productionize solution:

- Machine Learning Model as service, create Docker Image.
- Create ETL Pipeline to feed and periodically retrain ML model.
- Create tests and User Interface.