This score was built by SSC Analytics using Scoring Gizmo package:

Name of the Score

Name of the Author, Department of the Author

Description of the study

Entity name	
Study name	
Application	
Author	
Sponsor	
Recipients	
For information	
Operational Validation	
Domain / Field	
Train session	TRAIN_bg_stage2_2024-06-20 14:27:11.793896_no_tag

Period of study	2017-01 – 2020-03
Period of train	2017-01 – 2019-12 (70%)
Period of test	2017-01 – 2019-12 (30%)
Period of temporal validation	2020-01, 2020-02, 2020-03
Excluded periods	[]
Data origin	
Statistical tools	**Error! Bookmark not defined.**Python, Sklearn(Random Forest, Decision tree), Xgboost, Logistic regression
Structural validation method	Random split of building sample (70/30 ratio)
Criterion	
Performances	
Roles	

	AccuracyScore	AUC	PrecisionScore	Recall	F1	CRCR_pred_cutoff	Volumes	DataSet	Method	NbFeatures
0	0.30	0.51	0.34	0.34	0.28	1.25	0.94	118955	train_X	xgb	100
1	0.30	0.51	0.34	0.34	0.28	1.25	0.94	58590	test_X	xgb	100
2	0.28	0.49	0.32	0.32	0.27	1.24	0.88	4773	t1df	xgb	100
3	0.28	0.50	0.33	0.33	0.28	1.25	0.88	5453	t2df	xgb	100
1	0.30	0.51	0.34	0.34	0.28	1.25	0.94	58590	test_X	xgb	100
2	0.28	0.49	0.32	0.32	0.27	1.24	0.88	4773	t1df	xgb	100
3	0.28	0.50	0.33	0.33	0.28	1.25	0.88	5453	t2df	xgb	100
4	0.30	0.50	0.33	0.34	0.28	1.26	0.93	5019	t3df	xgb	100

Content

1. Presentation of the situation

1.1 The Overview

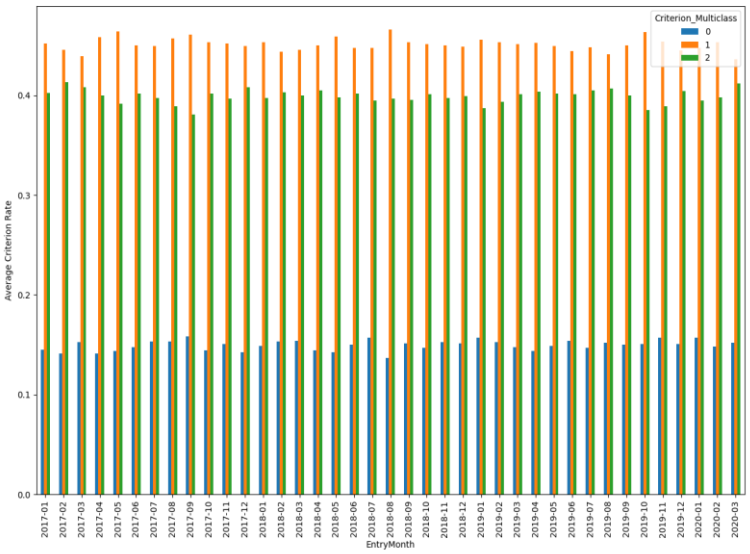
Fill in this part during the scoping definition of the project.

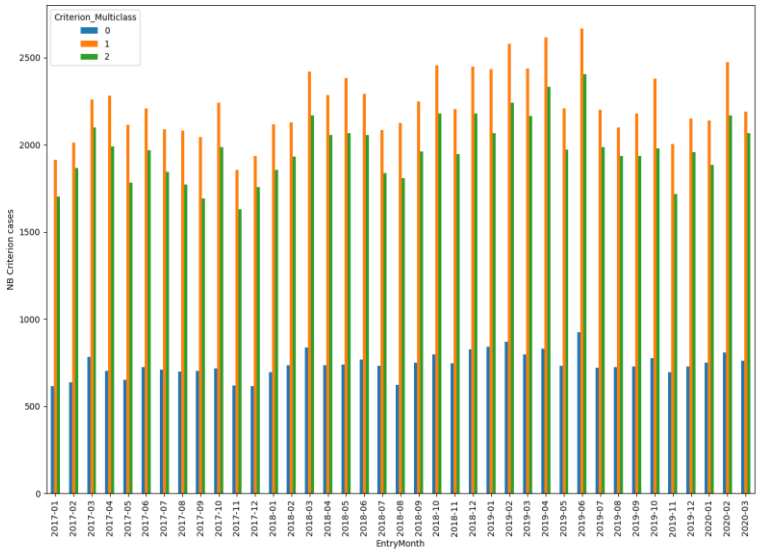
2. Presentation and analysis of project

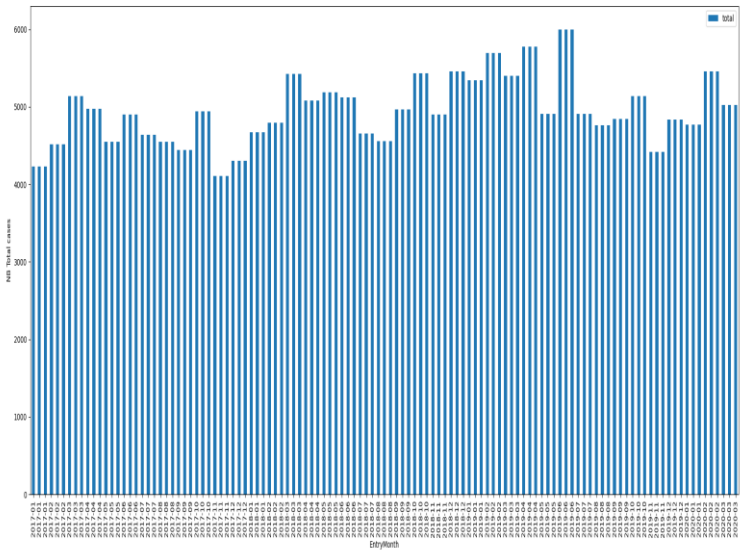
2.1 The perimeter of the study

Fill in this part during the scoping definition of the project.

2.1.1 Criterion study – the whole data set before removing periods

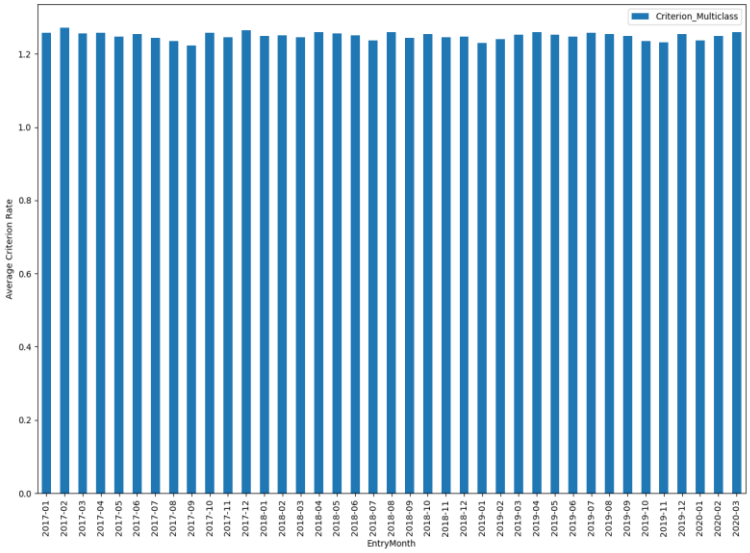


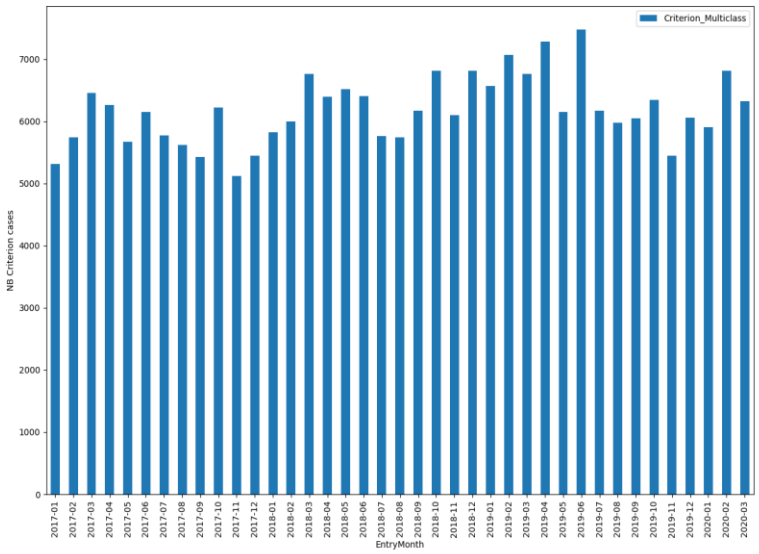


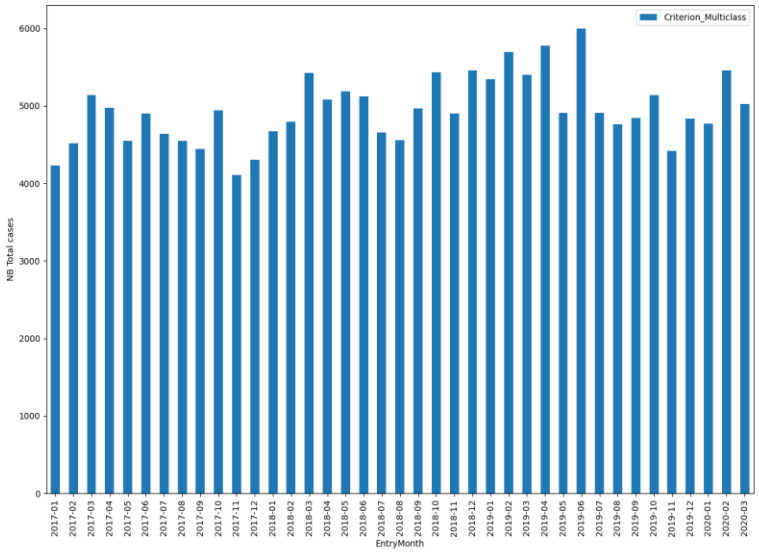


Insert your comments here

2.1.2 Criterion study – the after removing periods







Insert your comments here

2.2 The wherewithal

Dataset was built in SQL. The study done in Python.

2.3 Target variable

Fill in during the scoping phase.

3. Modelling

3.1 Statistical method

For the modelling part the dataset was split on 2 parts- Train (70%) and Test (30%) + 3 temporal validation samples.

The volumes are below:

- Train: 118955 nb cases, with 1.251 criterion rate
- Test: 58590 nb cases, with 1.248 criterion rate
- t1: 4773 nb cases, with 1.238 criterion rate
- t2: 5453 nb cases, with 1.25 criterion rate
- t3: 5019 nb cases, with 1.26 criterion rate

3.2 Candidate variables

Describe the source of the data.

The description of the dataset is in appendix

3.3 Model selection

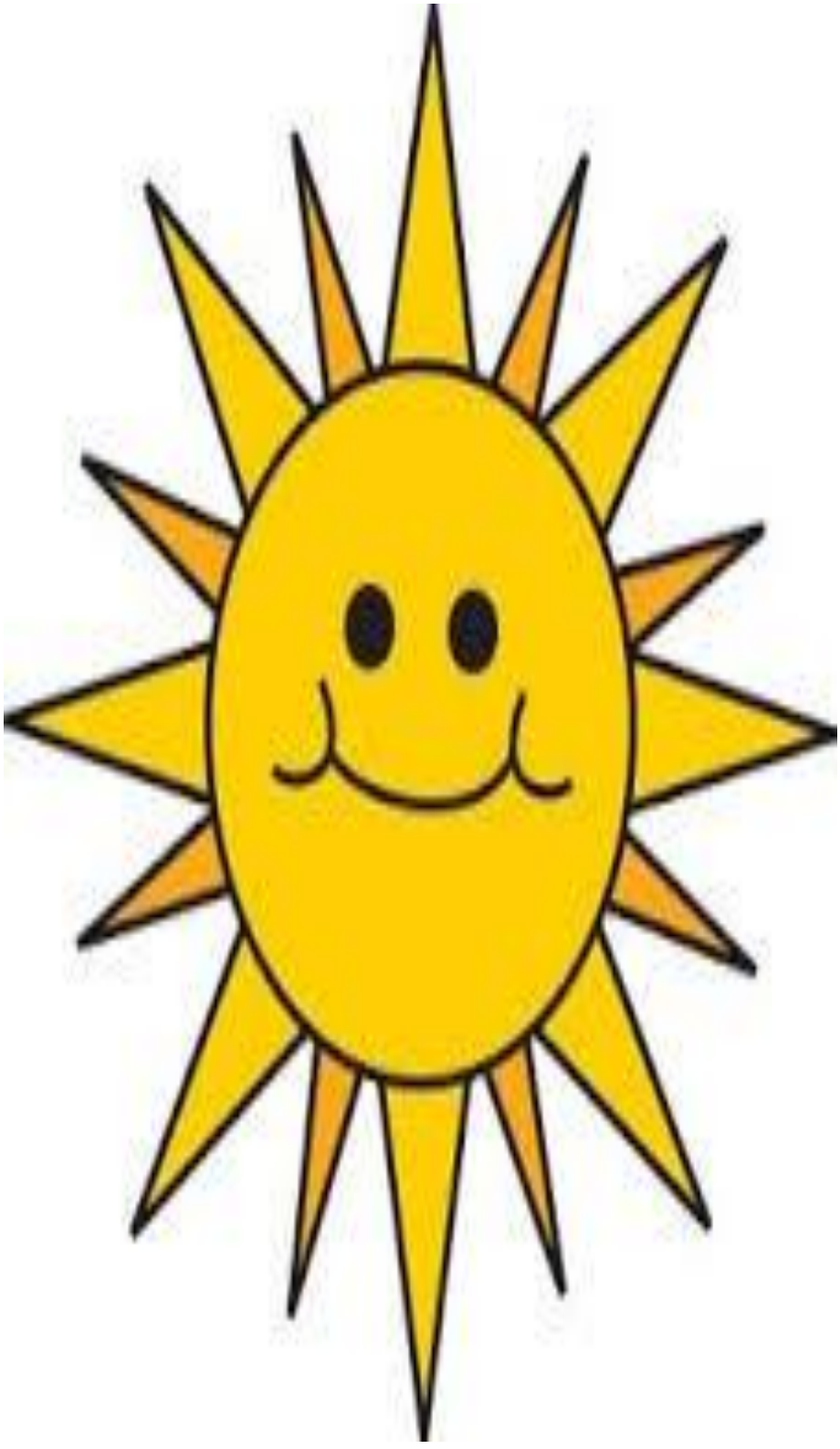
Couple of different statistical models were used in competition to be able to pick the best option from (1) efficiency point of view, but also considering the technical and business implementations.
Models used in competition: XGBoost, Random forest, Logistic regression, Decision tree

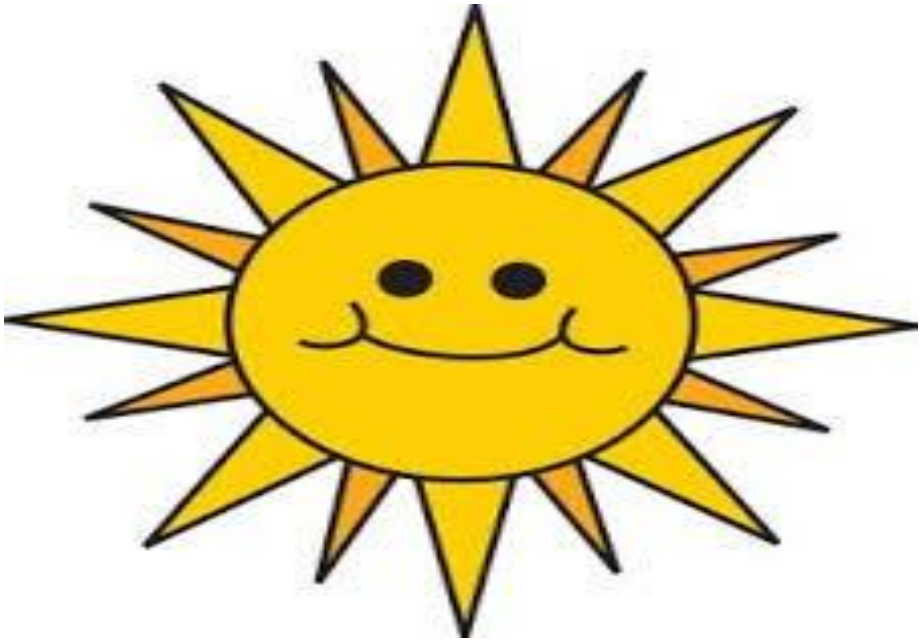
3.4 Variables selection

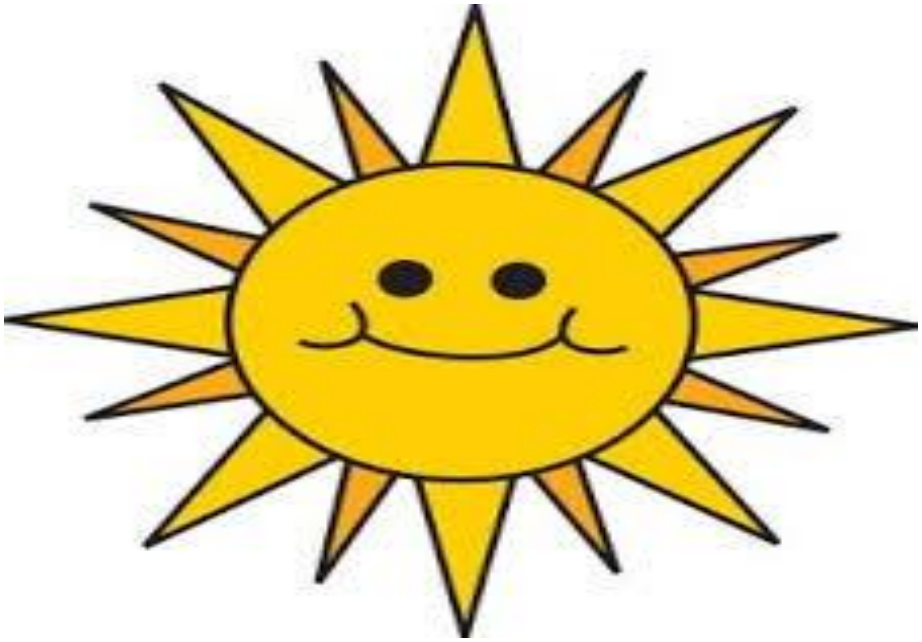
All variables were with treated missing files and after a feature engineering process was performed in order to enrich the predictors in the data set.
At the end, all features with correlation greater than 5% were used for modelling.
Share of missing values by variable is in the appendix.
Missing values were treated: 1

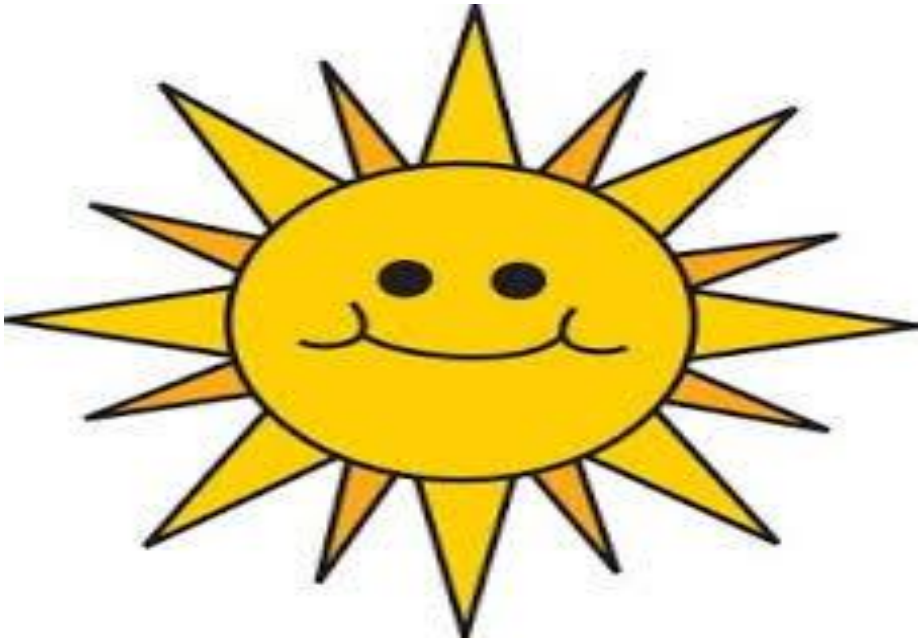
4. Final model

The chosen model is: XGBoost Classifier
Based on the nb of probabilities the following 2 bands (cut-offs) are possible:

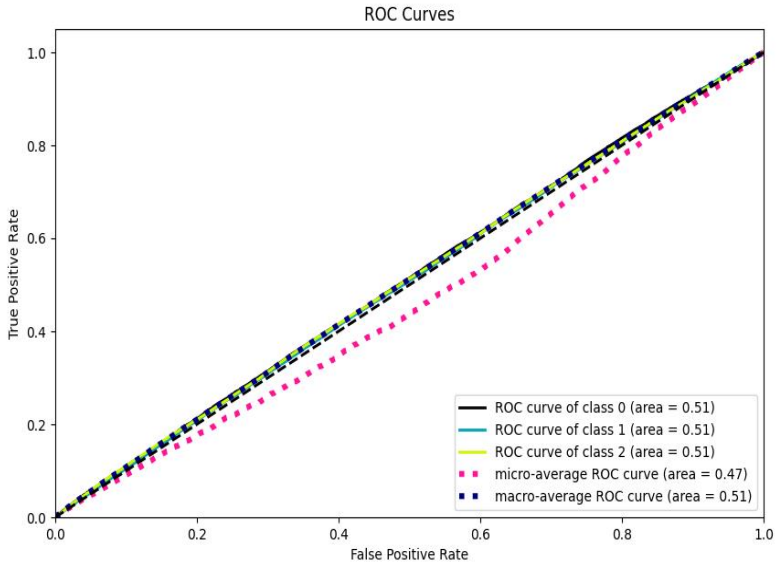




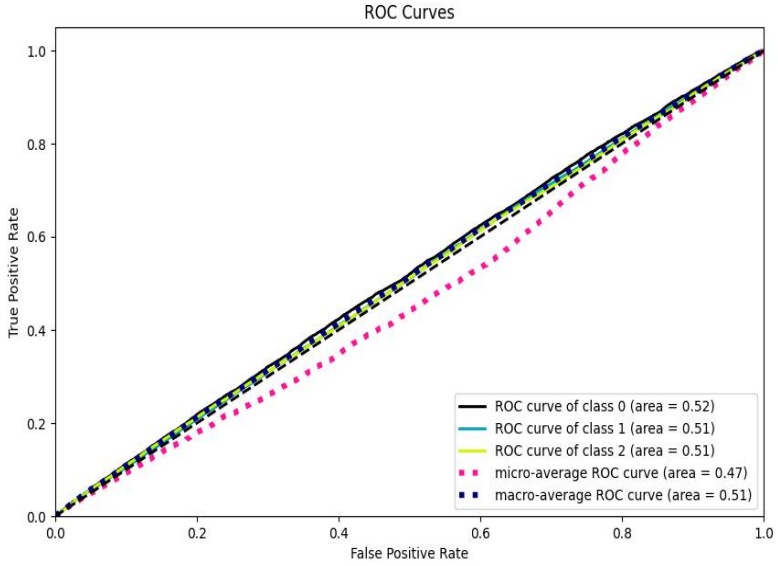




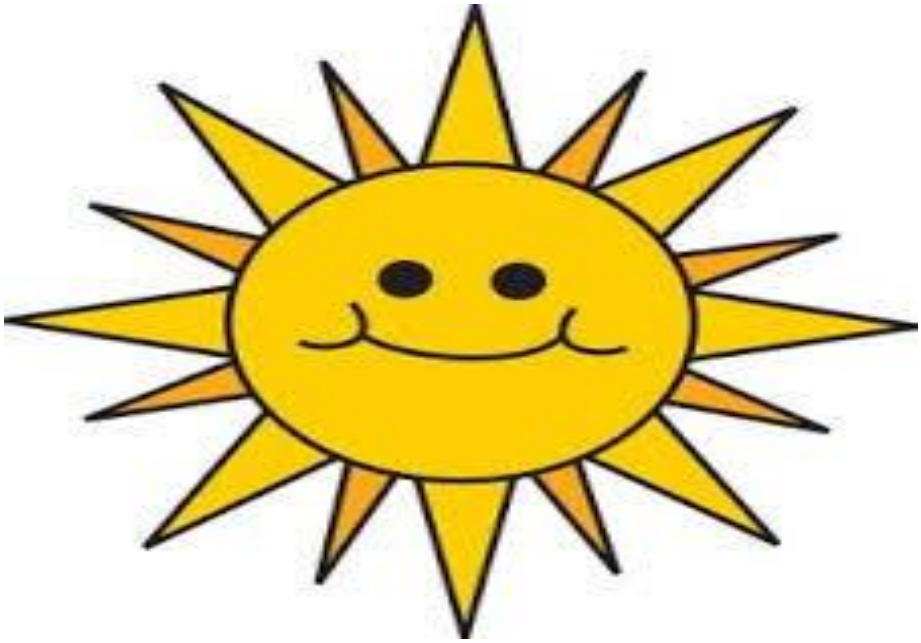
ROC Curve Train



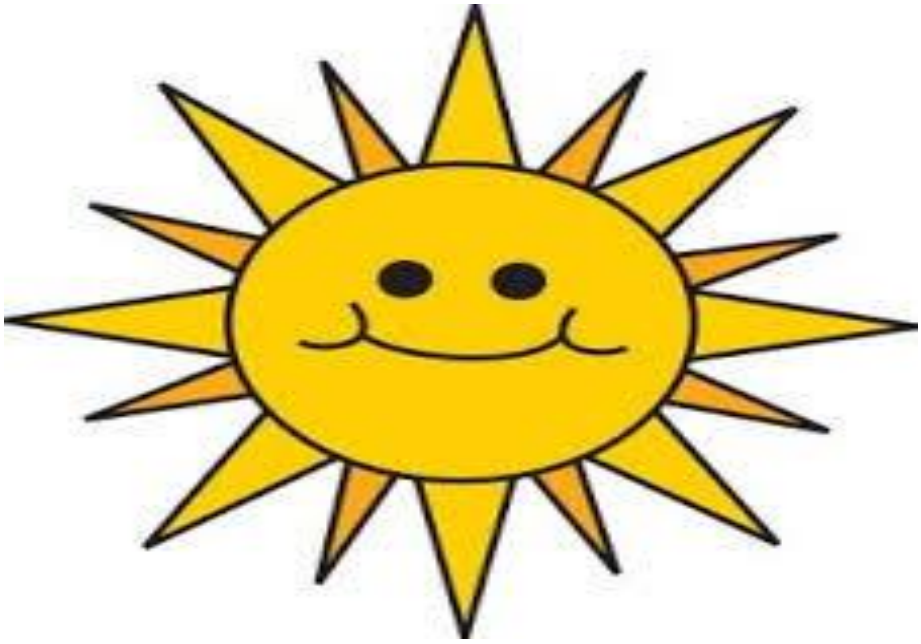
ROC Curve Test



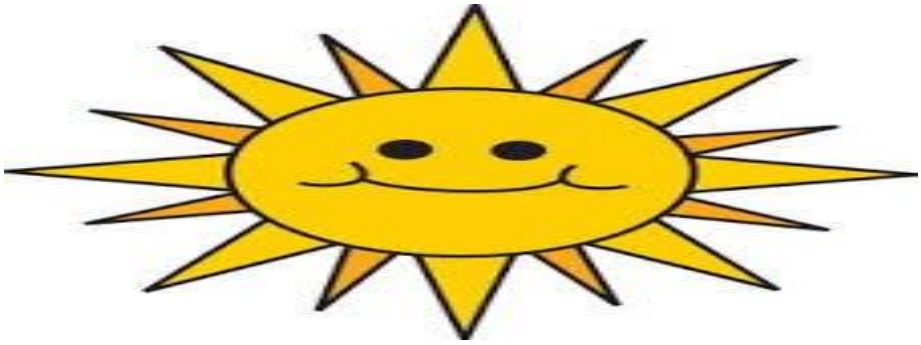
CAP curve for Train sample:



CAP curve for Test sample:

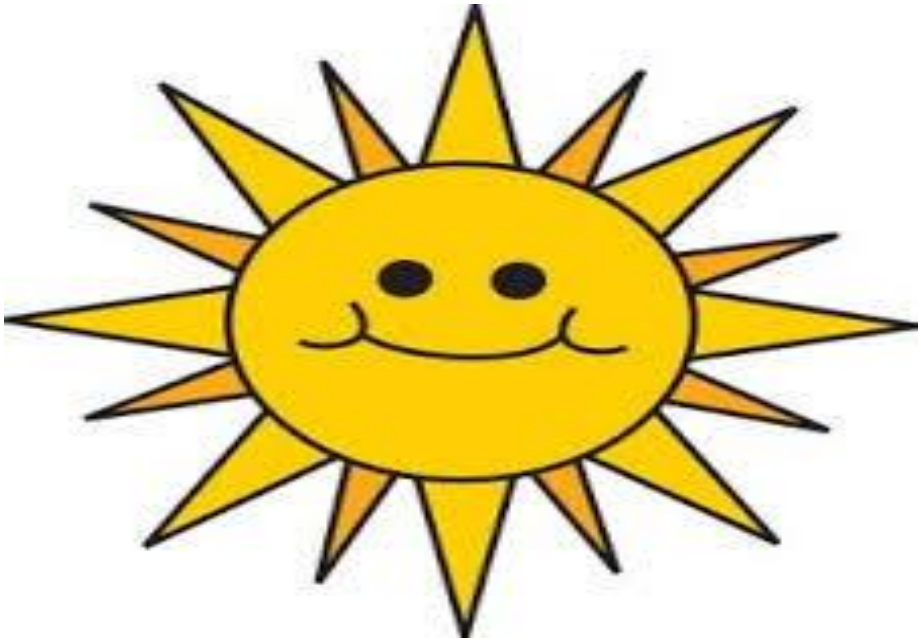


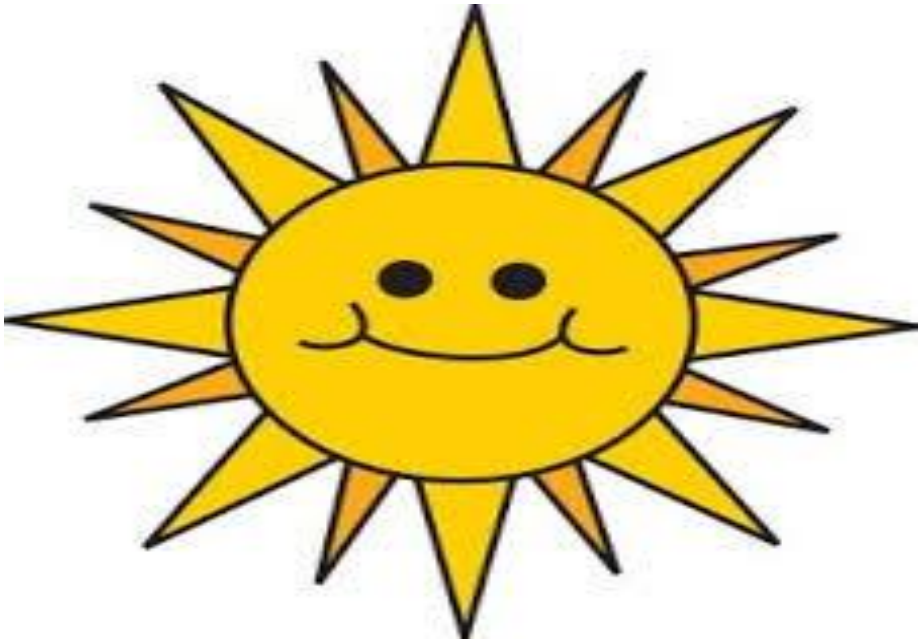
The split between bands for Train and Test with positive and negative criterion:

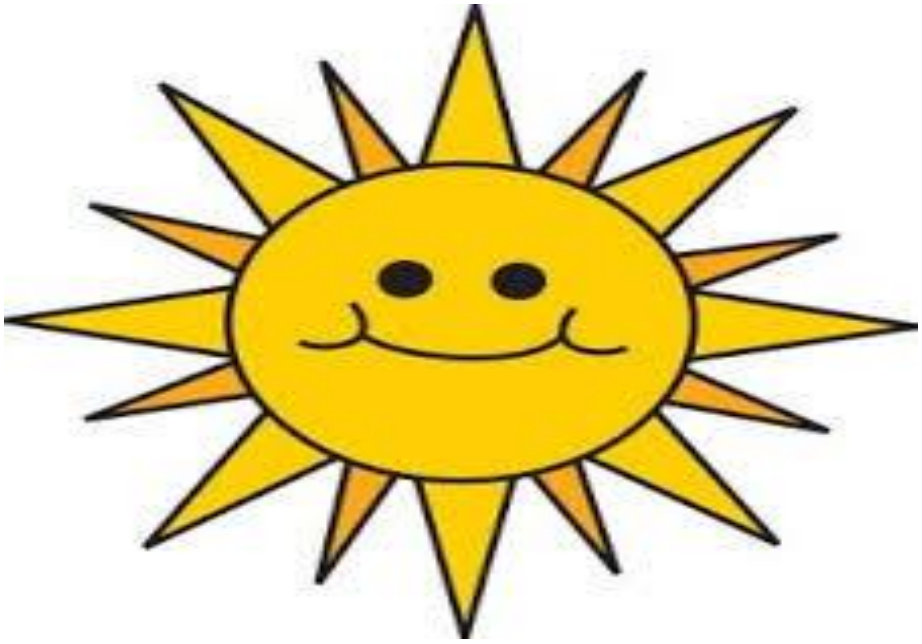


4.1 Segments stability

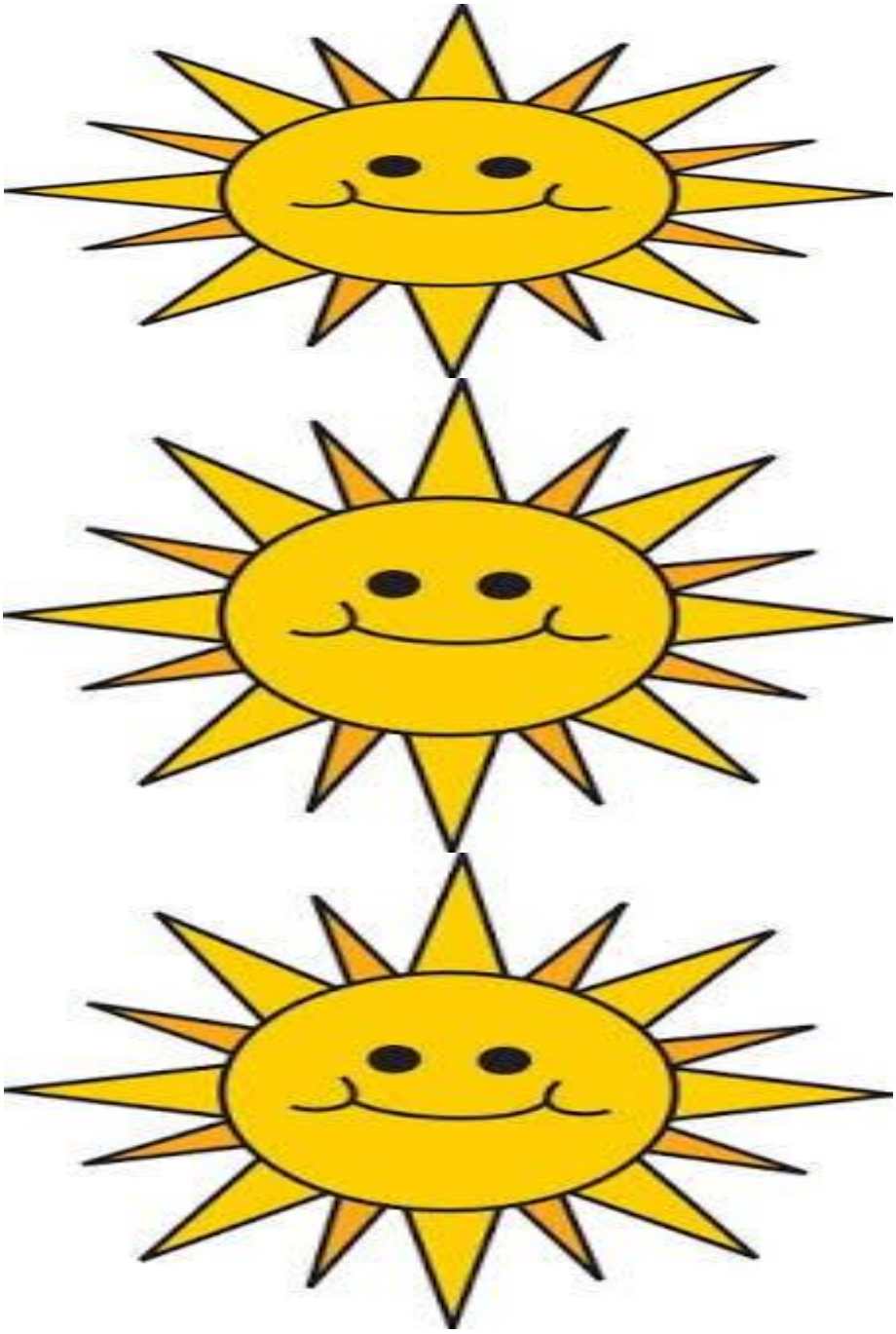
Evolution of Positive and Negative Criterion by bands and observation points for all periods in Train, Test, T1, T2 and T3:



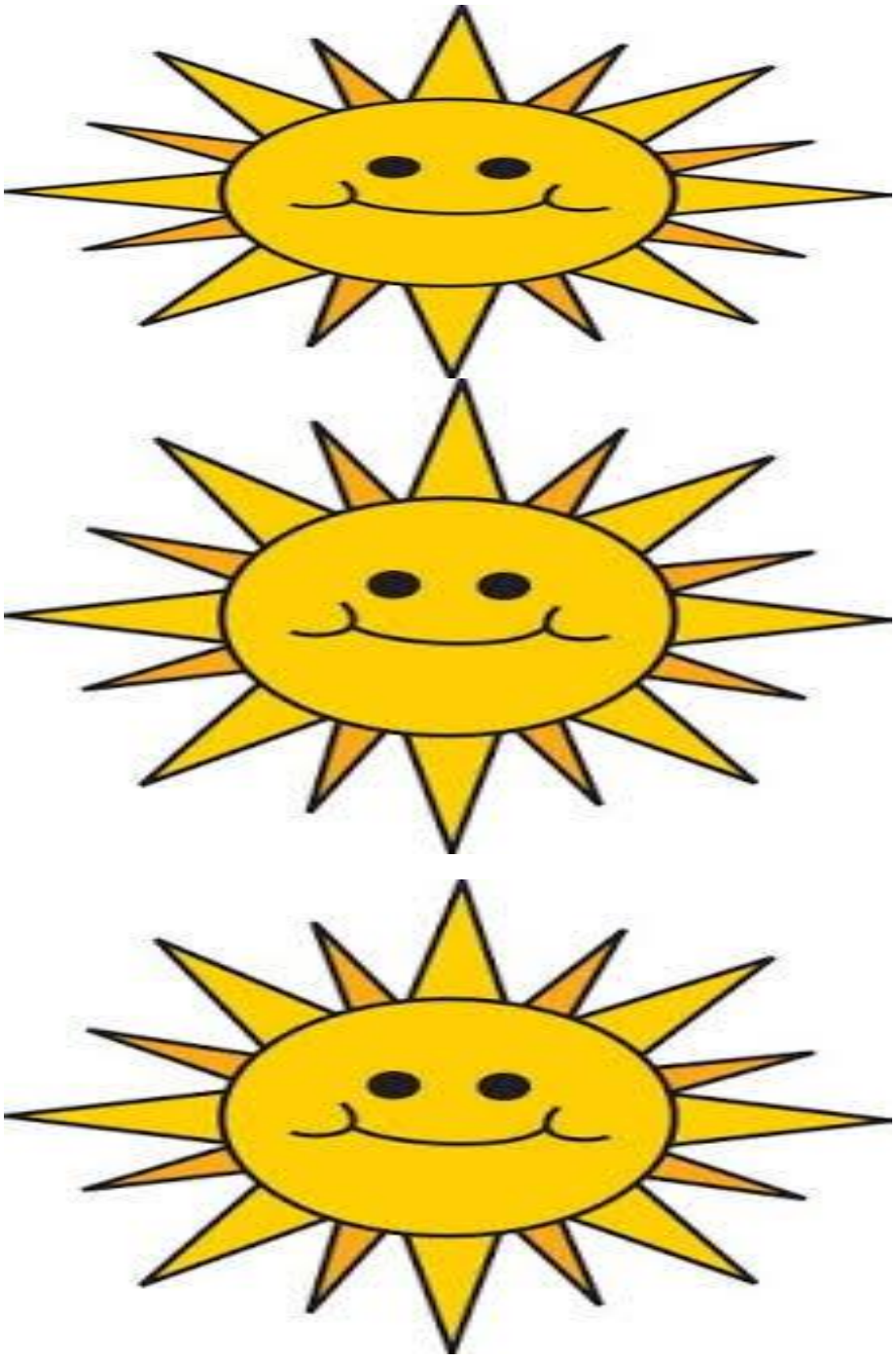




*For **Train** set in the graphs below are shown volumes the deciles, split by Proba, and the evolution of the criterion rate by decile:*



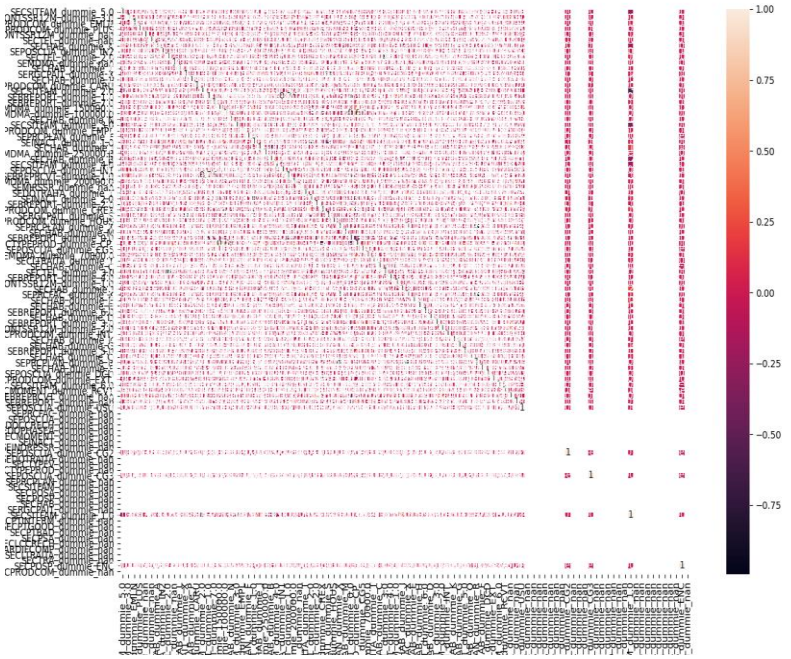
*For **Test** set in the graphs below are shown volumes the deciles, split by Proba, and the evolution of the criterion rate by decile:*

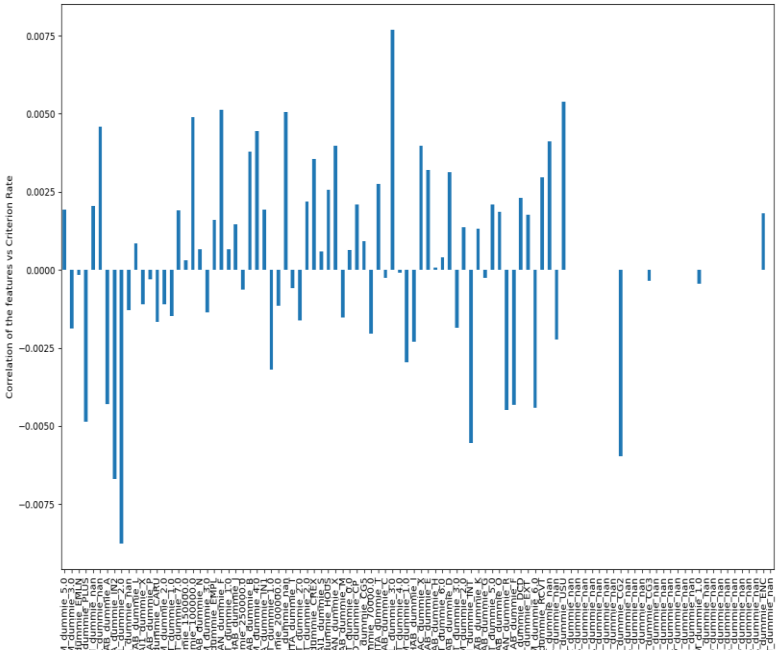


4.2 Variables stability

Pierson`s correlation between final and raw features used for the final model can be observed on the graph below.

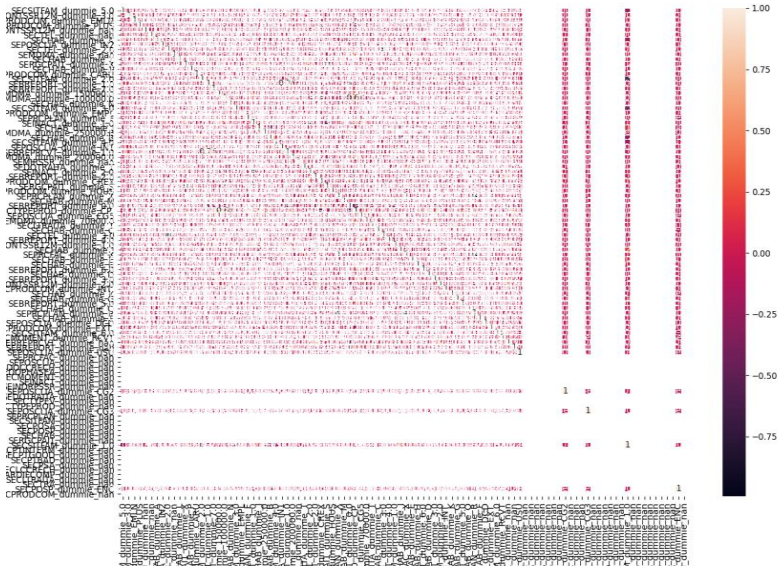
List of final features: **Error! Bookmark not defined.**['SECSITFAM_dummie_5.0' 'SEBMONTSSR12M_dummie_3.0' 'SECPRODCOM_dummie_EMLN' 'SECPRODCOM_dummie_PLUS' 'SEBMONTSSR12M_dummie_nan' 'SECTEL_dummie_nan' 'SECHAB_dummie_A' 'SEPOSCLIA_dummie_IN2' 'SECTEL_dummie_2.0' 'SEMDMA_dummie_nan' 'SECHAB_dummie_L' 'SERGCPAI1_dummie_X' 'SECHAB_dummie_P' 'SECPRODCOM_dummie_CARU' 'SECSITFAM_dummie_2.0' 'SEBREPORT_dummie_1.0' 'SEBREPORT_dummie_7.0' 'SEMDMA_dummie_150000.0' 'SEMDMA_dummie_100000.0' 'SECHAB_dummie_N' 'SECSITFAM_dummie_3.0' 'SECPRODCOM_dummie_EMPL' 'SEPRCPLAN_dummie_F' 'SEINACT_dummie_1.0' 'SECHAB_dummie_J' 'SEMDMA_dummie_250000.0' 'SECHAB_dummie_B' 'SECSITFAM_dummie_4.0' 'SEPOSCLIA_dummie_IN1' 'SEBREPRCVT_dummie_1.0' 'SEMDMA_dummie_200000.0' 'SEMIRSSR_dummie_nan' 'SEDOTRAITA_dummie_T' 'SEINACT_dummie_2.0' 'SEBREPORT_dummie_2.0' 'SECPRODCOM_dummie_CREX' 'SERGCPAI1_dummie_S' 'SECPRODCOM_dummie_HOUS' 'SEPRCPLAN_dummie_X' 'SECHAB_dummie_M' 'SEBREPORT_dummie_0.0' 'SECTYPEPROD_dummie_CP ' 'SEPOSCLIA_dummie_CG5' 'SEMDMA_dummie_70000.0' 'SECLTRAITA_dummie_T' 'SECHAB_dummie_C' 'SECTEL_dummie_3.0' 'SEBREPORT_dummie_4.0' 'SEBMONTSSR12M_dummie_1.0' 'SECHAB_dummie_I' 'SEPRCFAC_dummie_X' 'SECHAB_dummie_E' 'SECHAB_dummie_H' 'SEBREPORT_dummie_6.0' 'SECHAB_dummie_D' 'SEBREPORT_dummie_3.0' 'SEBMONTSSR12M_dummie_2.0' 'SECPRODCOM_dummie_INT ' 'SECHAB_dummie_K' 'SECHAB_dummie_G' 'SEBREPORT_dummie_5.0' 'SECHAB_dummie_O' 'SEPRCPLAN_dummie_R' 'SECHAB_dummie_F' 'SEPOSCLIA_dummie_DCD' 'SECPRODCOM_dummie_EXT ' 'SECSITFAM_dummie_6.0' 'SECMOMENT_dummie_RCVT' 'SEBREPRCVT_dummie_nan' 'SEBREPORT_dummie_nan' 'SEPOSCLIA_dummie_USU' 'SEPRCFAC_dummie_nan' 'SEPOSCLIA_dummie_nan' 'SEDOCCRECH_dummie_nan' 'SEDOPHASEA_dummie_nan' 'SECMOMENT_dummie_nan' 'SEINACT_dummie_nan' 'SEINDRPSSR_dummie_nan' 'SEPOSCLIA_dummie_CG2' 'SEDOTRAITA_dummie_nan' 'SECTYPEV_dummie_nan' 'SECTYPEPROD_dummie_nan' 'SEPOSCLIA_dummie_CG3' 'SEPRCPLAN_dummie_nan' 'SECSITFAM_dummie_nan' 'SECPOSA_dummie_nan' 'SECPOSP_dummie_nan' 'SECHAB_dummie_nan' 'SERGCPAI1_dummie_nan' 'SECSITFAM_dummie_1.0' 'SECPTINTERM_dummie_nan' 'SECPTGOOD_dummie_nan' 'SECPTBAD_dummie_nan' 'SECPSA_dummie_nan' 'SECLCCRECH_dummie_nan' 'SECARDIFCOMP_dummie_nan' 'SECLTRAITA_dummie_nan' 'SECTRA_dummie_nan' 'SECPOSP_dummie_ENC' 'SECPRODCOM_dummie_nan']





List of raw features: **Error! Bookmark not defined.**['SECSITFAM_dummie_5.0' 'SEBMONTSSR12M_dummie_3.0' 'SECPRODCOM_dummie_EMLN' 'SECPRODCOM_dummie_PLUS' 'SEBMONTSSR12M_dummie_nan' 'SECTEL_dummie_nan' 'SECHAB_dummie_A' 'SEPOSCLIA_dummie_IN2' 'SECTEL_dummie_2.0' 'SEMDMA_dummie_nan' 'SECHAB_dummie_L' 'SERGCPAI1_dummie_X' 'SECHAB_dummie_P'

'SECPRODCOM_dummie_CARU' 'SECSITFAM_dummie_2.0' 'SEBREPORT_dummie_1.0'
'SEBREPORT_dummie_7.0' 'SEMDMA_dummie_150000.0' 'SEMDMA_dummie_100000.0'
'SECHAB_dummie_N' 'SECSITFAM_dummie_3.0' 'SECPRODCOM_dummie_EMPL'
'SEPRCPLAN_dummie_F' 'SEINACT_dummie_1.0' 'SECHAB_dummie_J'
'SEMDMA_dummie_250000.0' 'SECHAB_dummie_B' 'SECSITFAM_dummie_4.0'
'SEPOSCLIA_dummie_IN1' 'SEBREPRCVT_dummie_1.0' 'SEMDMA_dummie_200000.0'
'SEMIRSSR_dummie_nan' 'SEDOTRAITA_dummie_T' 'SEINACT_dummie_2.0'
'SEBREPORT_dummie_2.0' 'SECPRODCOM_dummie_CREX' 'SERGCPAI1_dummie_S'
'SECPRODCOM_dummie_HOUS' 'SEPRCPLAN_dummie_X' 'SECHAB_dummie_M'
'SEBREPORT_dummie_0.0' 'SECTYPEPROD_dummie_CP ' 'SEPOSCLIA_dummie_CG5'
'SEMDMA_dummie_70000.0' 'SECLTRAITA_dummie_T' 'SECHAB_dummie_C'
'SECTEL_dummie_3.0' 'SEBREPORT_dummie_4.0' 'SEBMONTSSR12M_dummie_1.0'
'SECHAB_dummie_I' 'SEPRCFAC_dummie_X' 'SECHAB_dummie_E' 'SECHAB_dummie_H'
'SEBREPORT_dummie_6.0' 'SECHAB_dummie_D' 'SEBREPORT_dummie_3.0'
'SEBMONTSSR12M_dummie_2.0' 'SECPRODCOM_dummie_INT ' 'SECHAB_dummie_K'
'SECHAB_dummie_G' 'SEBREPORT_dummie_5.0' 'SECHAB_dummie_O'
'SEPRCPLAN_dummie_R' 'SECHAB_dummie_F' 'SEPOSCLIA_dummie_DCD'
'SECPRODCOM_dummie_EXT ' 'SECSITFAM_dummie_6.0' 'SECMOMENT_dummie_RCVT'
'SEBREPRCVT_dummie_nan' 'SEBREPORT_dummie_nan' 'SEPOSCLIA_dummie_USU'
'SEPRCFAC_dummie_nan' 'SEPOSCLIA_dummie_nan' 'SEDOCCRECH_dummie_nan'
'SEDOPHASEA_dummie_nan' 'SECMOMENT_dummie_nan' 'SEINACT_dummie_nan'
'SEINDRPSSR_dummie_nan' 'SEPOSCLIA_dummie_CG2' 'SEDOTRAITA_dummie_nan'
'SECTYPEV_dummie_nan' 'SECTYPEPROD_dummie_nan' 'SEPOSCLIA_dummie_CG3'
'SEPRCPLAN_dummie_nan' 'SECSITFAM_dummie_nan' 'SECPOSA_dummie_nan'
'SECPOSP_dummie_nan' 'SECHAB_dummie_nan' 'SERGCPAI1_dummie_nan'
'SECSITFAM_dummie_1.0' 'SECPTINTERM_dummie_nan' 'SECPTGOOD_dummie_nan'
'SECPTBAD_dummie_nan' 'SECPSA_dummie_nan' 'SECLCCRECH_dummie_nan'
'SECARDIFCOMP_dummie_nan' 'SECLTRAITA_dummie_nan' 'SECTRA_dummie_nan'
'SECPOSP_dummie_ENC' 'SECPRODCOM_dummie_nan']



1. Implementation of the score

1.1 Principle

To be decided

1.2 Description of the implementation strategy

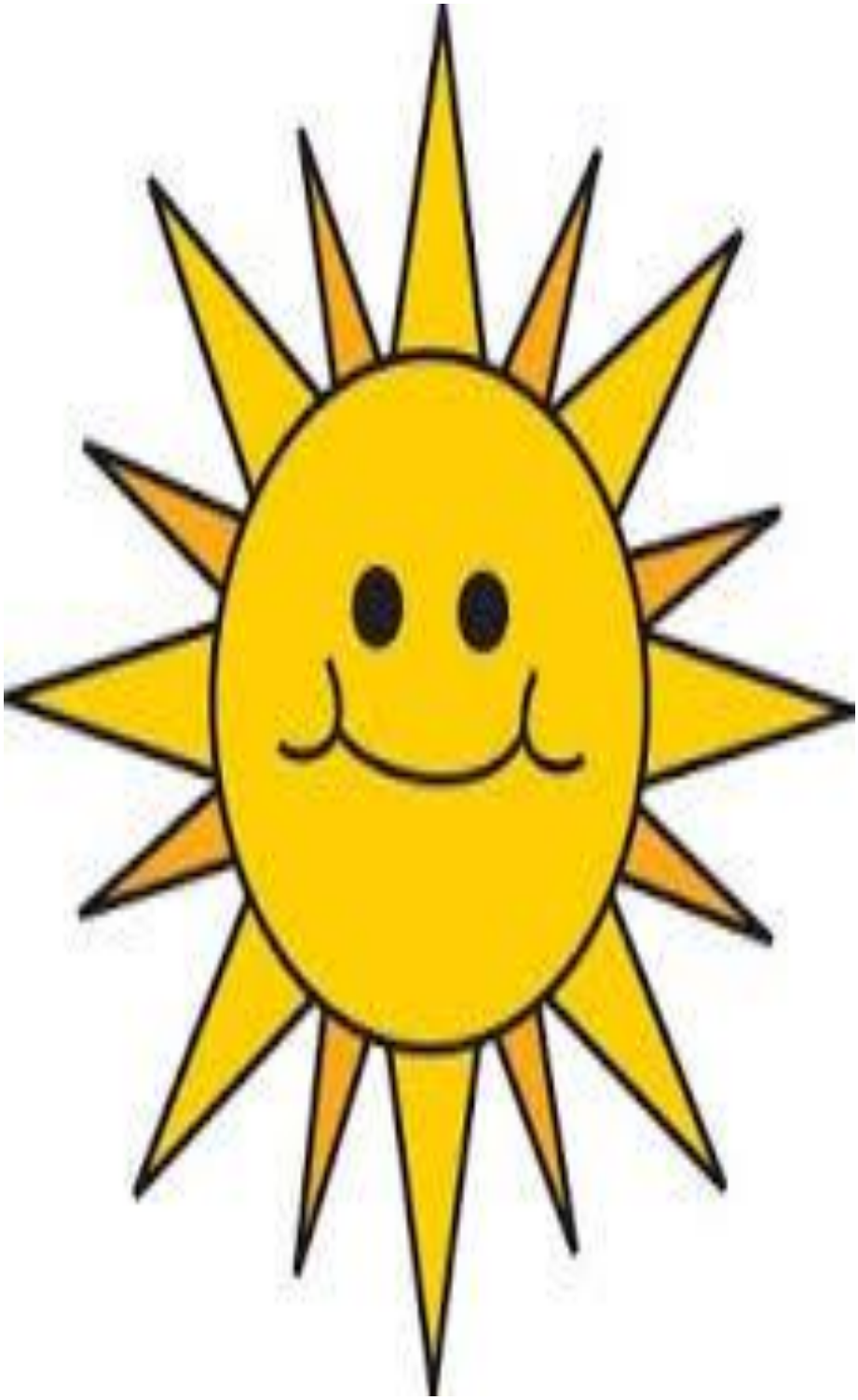
To be done

1.3 Follow up in production & real efficiency measures

MUSE to be implemented

2. Annex

2.1 Description of the data



2.2 Missing values

	column_name	percent_missing
5	Outstanding	2.198247
6	fkCredit	0.000519
9	ExitDate	25.620105
10	ES_Event	25.620105
13	LastPaymentDate	25.620105
14	SRFACBASE	25.620105
17	Litigation	63.413040
18	LitigationDate	63.413040
19	DPD180	98.290886
20	Dpd_Date	98.290886
21	CriterionDate	63.057213
22	CriterionOutst	63.057213
89	SEANCPROF	3.320193
90	SEBMENS	1.215831
91	SEBMONTSSR	1.958608
92	SEBMONTSSR12M	0.912910
93	SEBREPORT	3.253281
94	SEBREPRCVT	4.004876
95	SEBRGEAT	2.211733
111	SECPTT	0.064319
114	SECTEL	3.288552
127	SEDECOUVERT	2.573785
143	SEMCRD	2.208621
144	SEMDMA	2.080502
145	SEMDREG	0.988640
146	SEMIR	2.590902
147	SEMIRSSR	1.920224
148	SEMMENS	1.738679
149	SEMMENSP	1.726749
150	SEMREGORIG	2.241299
151	SEMREGSSR	2.575341
152	SEMREPENR	2.375642
153	SEMREPORTE	2.403652
154	SEMSREC	1.948234
155	SEMSRECAGIOS	2.255304
156	SEMSRECCAP	1.819078
158	SEMSRECDIVERS	0.640075
159	SEMSRECRET	1.921261
160	SEMTME	2.311323
163	SENBDOS	0.965818
164	SENBPLAN	0.136937
166	SEPHASE	0.367239
174	SEPHNBORIG___2	1.764096
175	SEPHNBORIG___4	2.255822
176	SEPHNBORIG___7	2.210696
177	SEPHNBORIG	2.133409
186	SETODUCLT	2.438404
187	SETODUDOSS	2.198247
215	SEMAXDPDEVER	89.585041
216	SECURRENTDPD	89.585041
217	SEPHNBORIG___3	7.907568
218	MM_EntryDate_LastPayment	25.620105
220	MM_SEDDREG_EntryDate	5.743036
6	fkCredit	0.000519
9	ExitDate	25.620105
10	ES_Event	25.620105
13	LastPaymentDate	25.620105
14	SRFACBASE	25.620105
17	Litigation	63.413040
18	LitigationDate	63.413040
19	DPD180	98.290886
20	Dpd_Date	98.290886
21	CriterionDate	63.057213
22	CriterionOutst	63.057213
89	SEANCPROF	3.320193
90	SEBMENS	1.215831
91	SEBMONTSSR	1.958608
92	SEBMONTSSR12M	0.912910
93	SEBREPORT	3.253281
94	SEBREPRCVT	4.004876
95	SEBRGEAT	2.211733
111	SECPTT	0.064319
114	SECTEL	3.288552
127	SEDECOUVERT	2.573785
143	SEMCRD	2.208621
144	SEMDMA	2.080502
145	SEMDREG	0.988640
146	SEMIR	2.590902
147	SEMIRSSR	1.920224
148	SEMMENS	1.738679
149	SEMMENSP	1.726749
150	SEMREGORIG	2.241299
151	SEMREGSSR	2.575341
152	SEMREPENR	2.375642
153	SEMREPORTE	2.403652
154	SEMSREC	1.948234
155	SEMSRECAGIOS	2.255304
156	SEMSRECCAP	1.819078
158	SEMSRECDIVERS	0.640075
159	SEMSRECRET	1.921261
160	SEMTME	2.311323
163	SENBDOS	0.965818
164	SENBPLAN	0.136937
166	SEPHASE	0.367239
174	SEPHNBORIG___2	1.764096
175	SEPHNBORIG___4	2.255822
176	SEPHNBORIG___7	2.210696
177	SEPHNBORIG	2.133409
186	SETODUCLT	2.438404
187	SETODUDOSS	2.198247
215	SEMAXDPDEVER	89.585041
216	SECURRENTDPD	89.585041
217	SEPHNBORIG___3	7.907568
218	MM_EntryDate_LastPayment	25.620105
220	MM_SEDDREG_EntryDate	5.743036
223	MM_SEDFIN_EntryDate	1.472587

2.3 Feature stability