

RISK MANAGEMENT PROJECT

Presented by:
Alessandro Bachetti
Silvia Palanca



01 WHAT IS THE GOAL

Project Management Risk

This study is conducted upon a dataset containing operational, human, technical, organizational, and external **risk factors** that influence **project** performance and **outcomes**. Our goal is to demonstrate how statistical methods can support early risk identification and decision-making in project management.

Specifics

- 4000 **projects** observed
- 50 **features** (25 numerical and 25 categorical)
- **Target variable:** RISK LEVEL
- 4 **levels:** Low, Medium, High, Critical



02 DATA IMPUTATION

Numerical

We conducted a check for every entry of the dataset looking for **na**'s in the numerical variables.

We then imputed the **median** for every missing value eventually found and **scaled** the variables for preventing imbalances within the dataset.

Categorical

We conducted the same analysis for categorical variables. We then imputed the **mode** for every missing value.



03 FEATURE SELECTION

Numerical variables 1

Correlation Matrix a priori

By performing a correlation analysis a priori we filtered the correlated variables ($\rho > 0.7$) and dropped them to prevent noise (multicollinearity) in the analysis.

Dropped features:

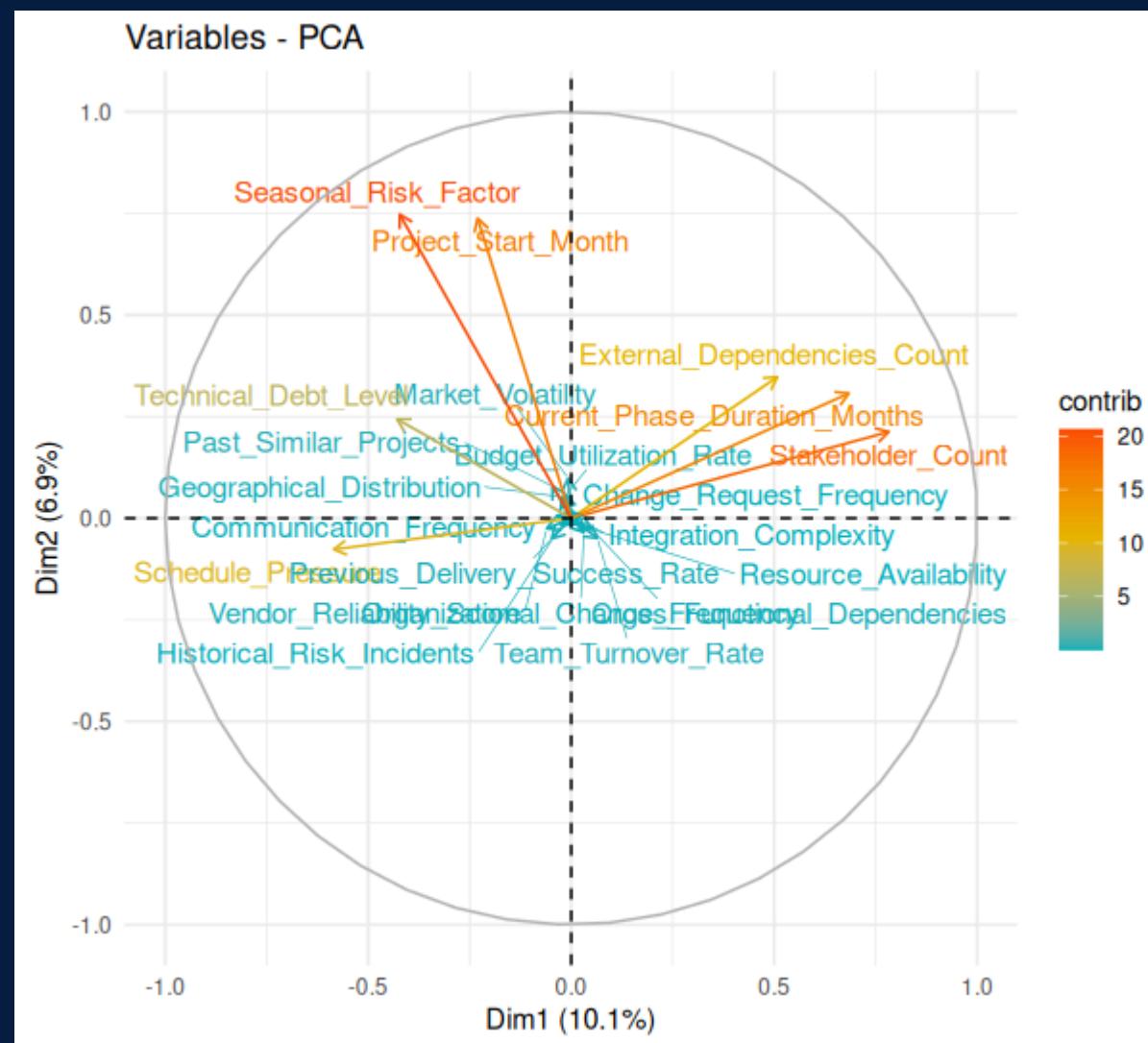
- [1] "Estimated_Timeline_Months"
- [2] "Project_Budget_USD"
- [3] "Team_Size"
- [4] "Complexity_Score"



03 FEATURE SELECTION

Numerical variables 2

PCA and ANOVA

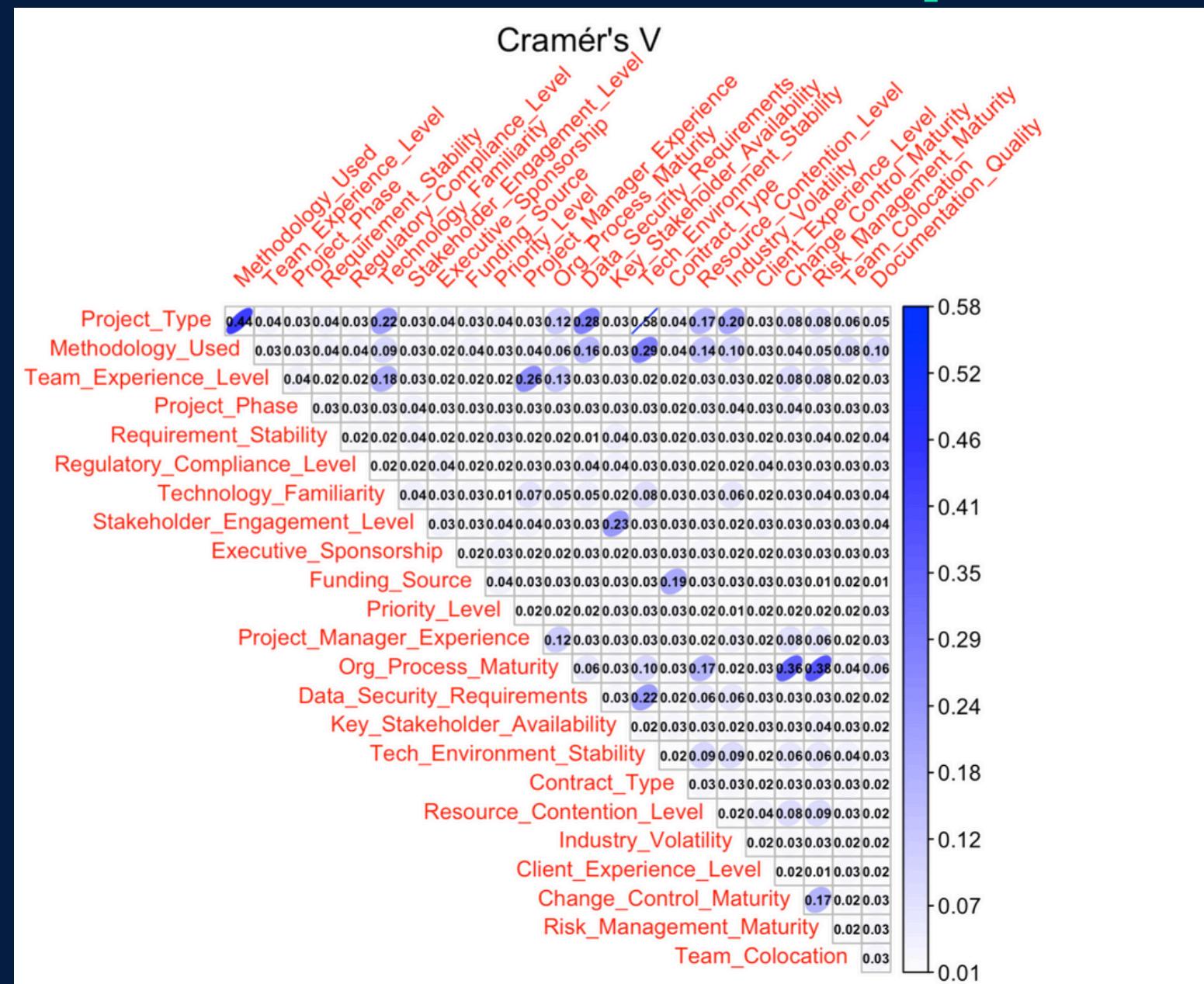


We then used a combination of PCA and ANOVA to select the variables that were most significant in explaining the variance of the data (top 5 variables that give the biggest contributions to the PC1 and PC2) and that were more statistically relevant (smallest p-values in ANOVA table).

03 FEATURE SELECTION

Categorical Variables

Cramér's V and Chi-squared



The selection of the categorical variables has been conducted via a Cramér's V analysis a priori (removed correlated categorical variables) and then by selecting the features with the Chi-squared criteria (top 5 most correlated to the target variable).

03 FEATURE SELECTION

- | | |
|--|---|
| <p>► EXTERNAL DEPENDENCIES COUNT Number of external factors or entities the project relies on.
Mean: 3.13
Std. Deviation: 1.61</p> <p>► TEAM TURNOVER RATE Percentage of team members leaving the project during its lifecycle.
Mean: 0.29
Std. Deviation: 0.17</p> <p>► CURRENT PHASE DURATION MONTHS The duration (in months) spent in the current project phase.
Mean: 4.07
Std. Deviation: 2.88</p> <p>► STAKEHOLDER COUNT The number of key individuals or groups with an interest in the project.
Mean: 11.1
Std. Deviation: 4.43</p> <p>► PAST SIMILAR PROJECTS Number of similar projects previously completed by the organization or team.
Mean: 1.97
Std. Deviation: 1.75</p> <p>► SEASONAL RISK FACTOR A multiplier indicating increased risk due to seasonal influences in certain industries.
Mean: 1.01
Std. Deviation: 0.03</p> | <p>► PROJECT START MONTH The numerical month (1-12) in which the project officially started.
Mean: 6.49
Std. Deviation: 3.48</p> <p>► ORG PROCESS MATURITY The maturity level of the organization's project management processes and standards.<ul style="list-style-type: none">• Levels: Optimising, Defined, Managed, Ad-hoc</p> <p>► TEAM EXPERIENCE LEVEL Overall experience level of the project team.<ul style="list-style-type: none">• Levels: Junior, Senior, Expert, Mixed</p> <p>► TECHNOLOGY FAMILIARITY How familiar the project team is with the core technologies used in the project.<ul style="list-style-type: none">• Levels: New, Familiar, Expert</p> <p>► PROJECT MANAGER EXPERIENCE The experience level of the assigned Project Manager.<ul style="list-style-type: none">• Levels: Junior, Mid-level, Senior, Certified</p> <p>► INDUSTRY VOLATILITY How stable or unpredictable the industry sector in which the project operates is.<ul style="list-style-type: none">• Levels: Stable, Moderate, High, Extreme</p> |
|--|---|

04

MULTINOMIAL MODELS



LOGIT (4 LEVELS)

By taking as reference level “low”, we conducted a multinomial logistic regression and checked with a confusion matrix whether the model predicted correctly the data or not



LDA

We then conducted a Linear Discriminant Analysis to double-check the results of the Multinomial Logistic Model.

04 MULTINOMIAL LOGISTIC REGRESSION

Confusion Matrix and Statistics

Reference						
		Prediction	Low	Medium	High	Critical
Low	103	53	8	2		
Medium	122	267	145	59		
High	13	73	89	60		
Critical	3	25	68	107		

Overall Statistics

Accuracy : 0.4728

Statistics by Class:

	Class: Low	Class: Medium	Class: High	Class: Critical
Sensitivity	0.42739	0.6388	0.28710	0.46930
Specificity	0.93410	0.5815	0.83540	0.90093

04 LINEAR DISCRIMINANT ANALYSIS

Confusion Matrix and Statistics

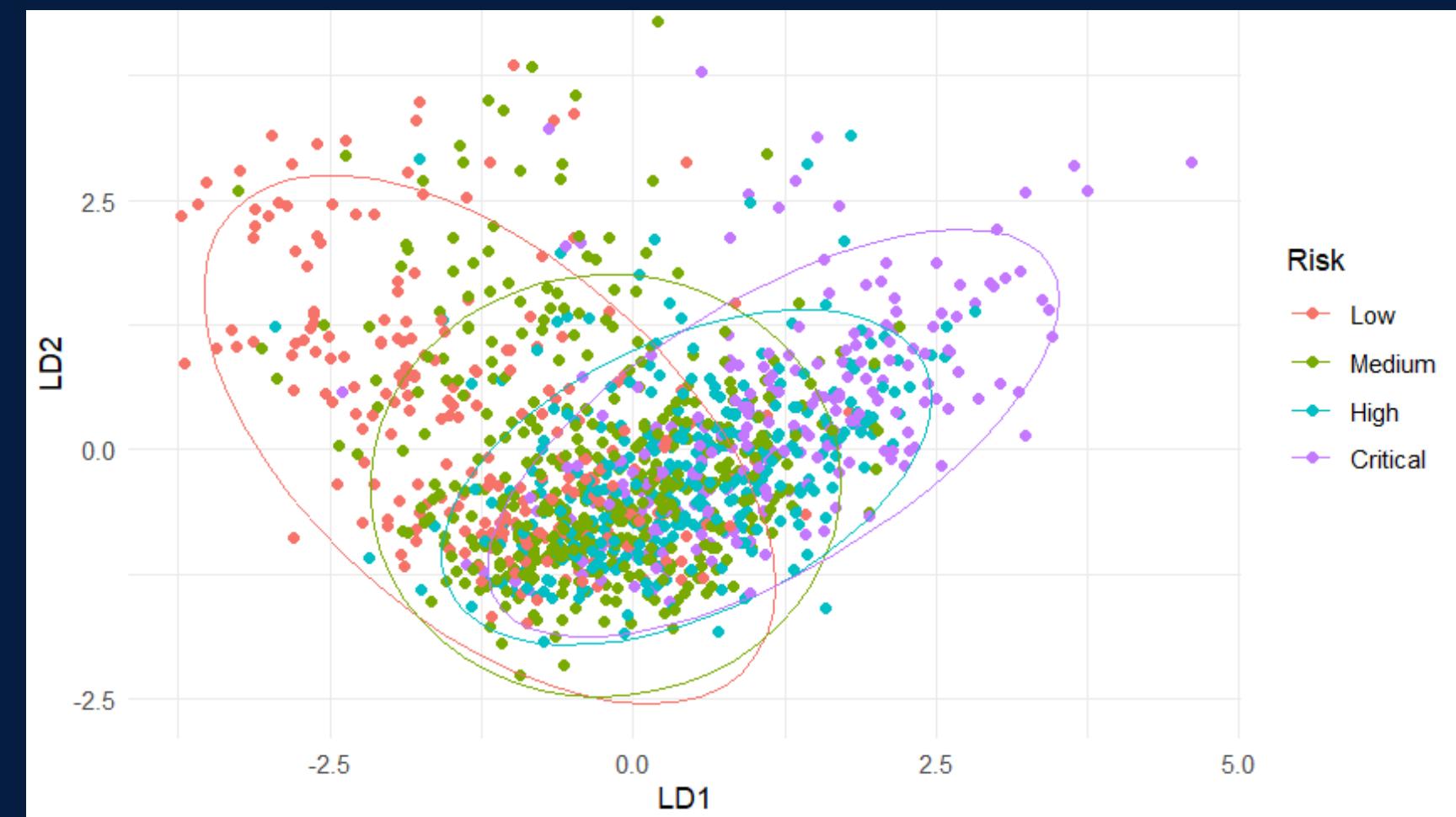
Reference					
Prediction	Low	Medium	High	Critical	
Low	105	61	9	4	
Medium	121	261	144	59	
High	10	69	92	52	
Critical	5	27	65	113	

Overall Statistics

Accuracy : 0.477

Statistics by Class:

	Class: Low	Class: Medium	Class: High	Class: Critical
Sensitivity	0.43568	0.6244	0.29677	0.4956
Specificity	0.92259	0.5841	0.85231	0.8999



05 FINAL APPROACH

Binary Model

To better understand, in Time-Critical scenarios, how to effectively support risk mitigation measures, we also decided to perform some binary analyses due to some confusion in the former Multinomial Models.

Categories “Low” and “Medium” are encoded as 0 and “High” and “Critical” as 1.

As we expected, the performance’s boost is relevant, therefore it represents a good support in decision making.

05 FINAL RESULTS

Confusion Matrix and Statistics Logit

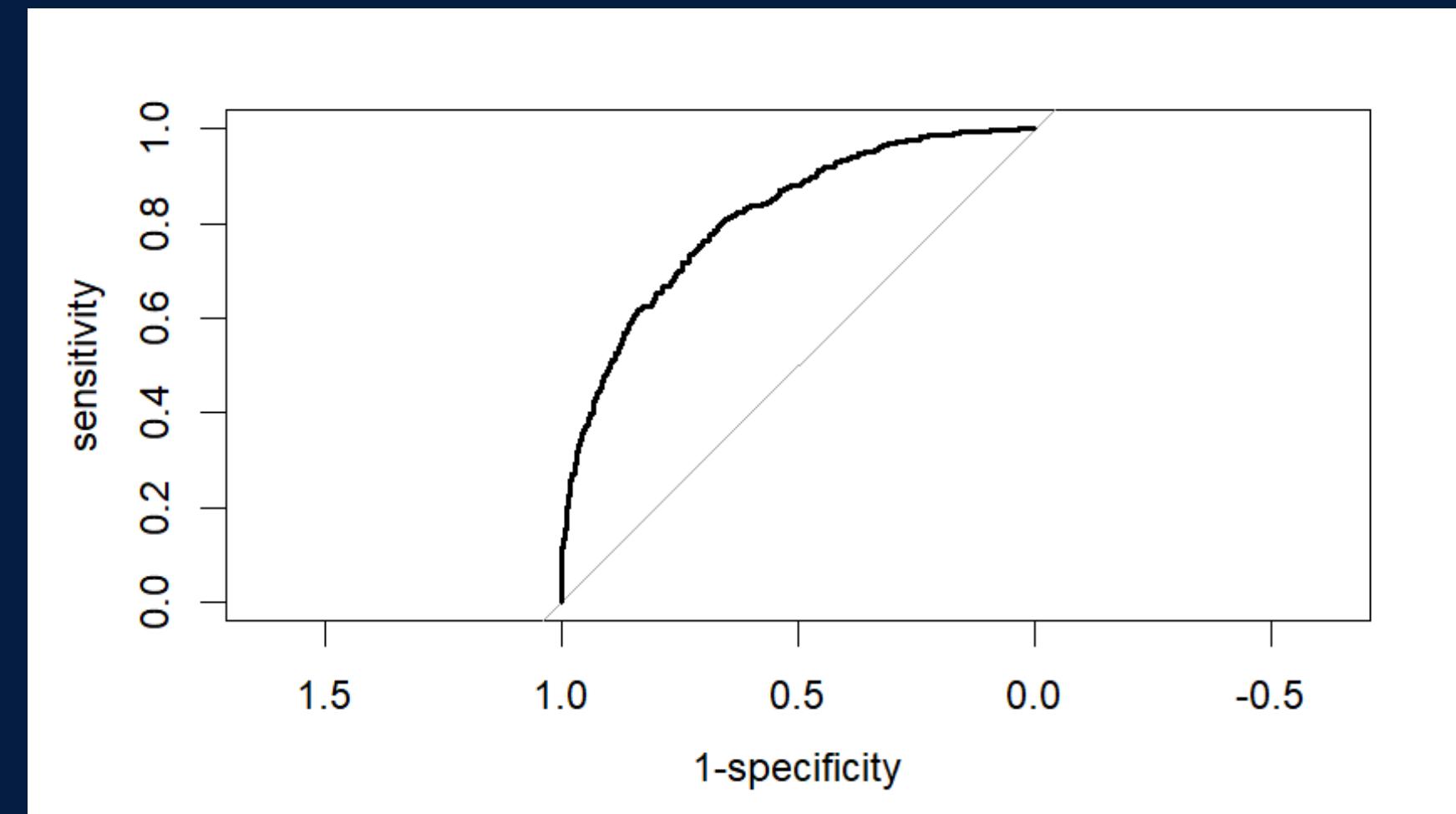
	Reference	
Prediction	0	1
0	515	179
1	144	359

Sensitivity : 0.7815
Specificity : 0.6673

Confusion Matrix and Statistics LDA

	Reference	
Prediction	0	1
0	521	183
1	138	355

Sensitivity : 0.7906
Specificity : 0.6599



AUC Binary Model: 0.811

06 CONCLUSIONS

	Modello	Accuracy
1	Logit Multinom	0.4728
2	LDA Multinom	0.4770
3	Logit Binary	0.7302
4	LDA Binary	0.7318



THANK YOU



alessandro.bachetti00@gmail.com



infopalancasilvia@gmail.com



https://github.com/alessandro-bachetti/risk_management_project

