



UNIVERSITÀ DI PISA

University of Pisa

Laurea Magistrale (MSc) in Artificial Intelligence and Data Engineering

Project

Data Mining and Machine Learning

Article Categorizer

Academic year 2020-2021

Lorenzo Bianchi

Github: <https://github.com/lorebianchi98/ArticlesCategorizer.git>



Introduction

Key idea

Formula 1, Max Verstappen su Red Bull vince il Gran premio di Austria. La Ferrari è solo quinta – Il video

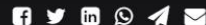
4 LUGLIO 2021 - 17:08

di Redazione



Il pilota olandese è arrivato al traguardo davanti a Valtteri Bottas (su Mercedes) e a Lando Norris (su McLaren)

INCORPORA VIDEO



È Max Verstappen il vincitore del Gran premio di Austria sul circuito Red bull ring di Spielberg. Il pilota olandese, su Red bull, ha condotto al primo posto tutta la gara, arrivando al traguardo davanti a Valtteri Bottas (su Mercedes) e a Lando Norris (su McLaren). Il quarto posto è invece di Lewis Hamilton (su Mercedes). Solo quinta la Ferrari guidata da Carlos Sainz. Così facendo, Verstappen allunga nella classifica generale piloti del campionato di Formula Uno. Ora ha 182 punti, 32 di vantaggio su Lewis Hamilton che oggi, appunto, non è andato oltre il quarto posto. Le immagini mostrano il pilota olandese mentre taglia il traguardo.

classifier

sport



OPEN



Sezioni

AMBIENTE

CULTURA & SPETTACOLO

~~EDITORIALI~~

~~LE NOSTRE STORIE~~

POLITICA

SPORT

ATTUALITÀ

ECONOMIA & LAVORO

~~FACT CHECKING~~

MONDO

SCIENZE

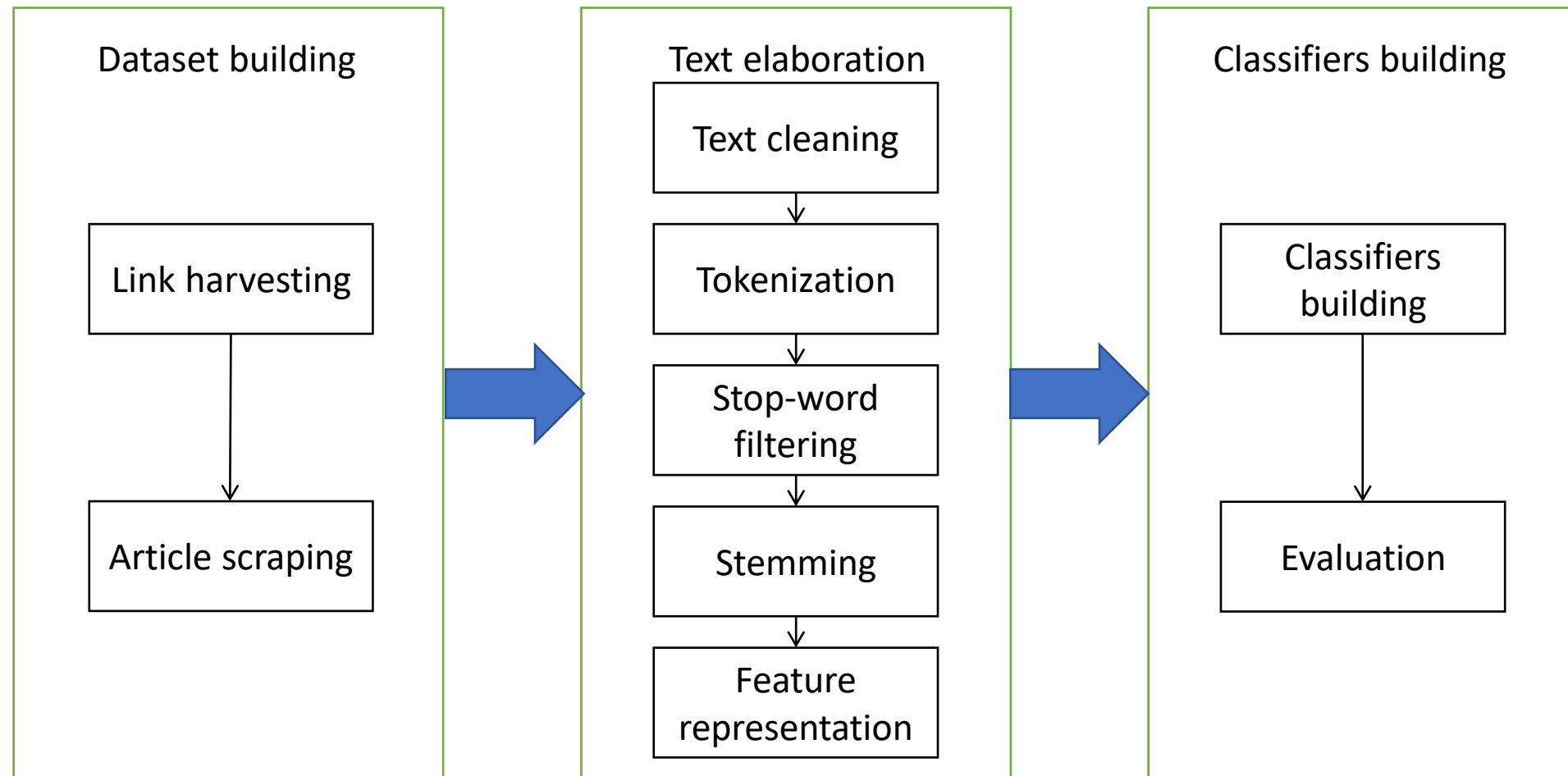
TECNOLOGIA

Categories

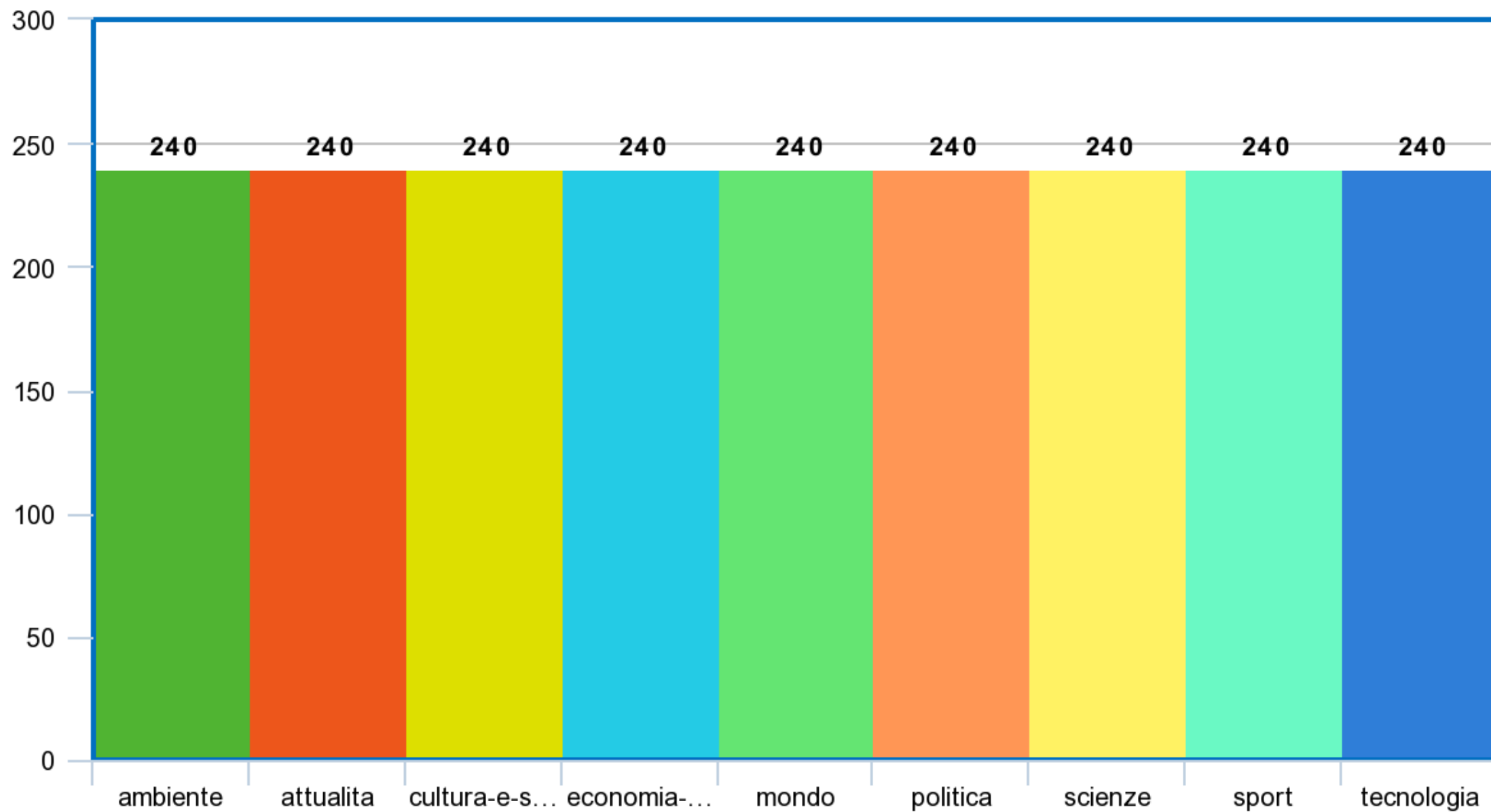


The application

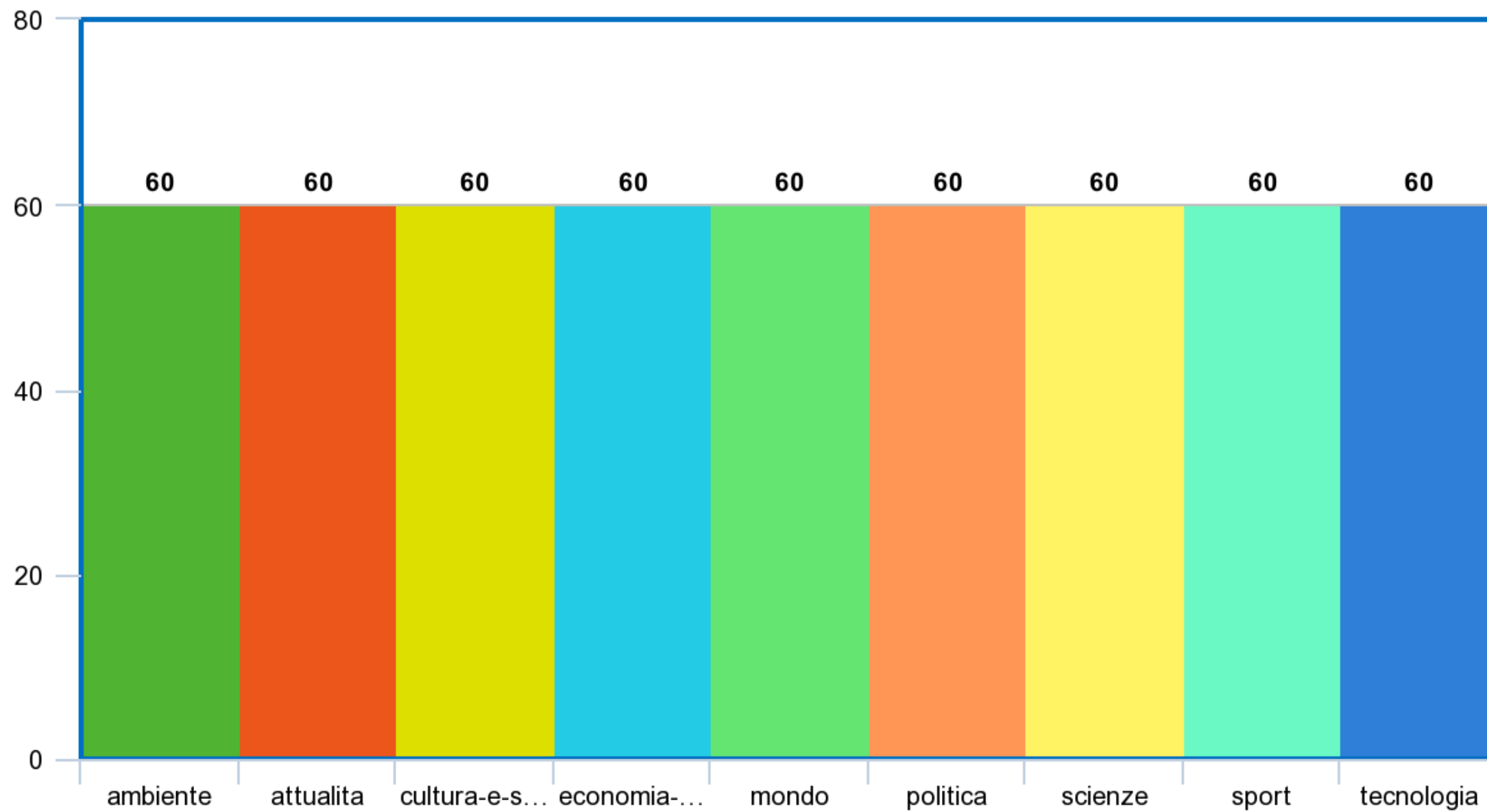
Phases of the elaboration



Training set distribution



Test set distribution

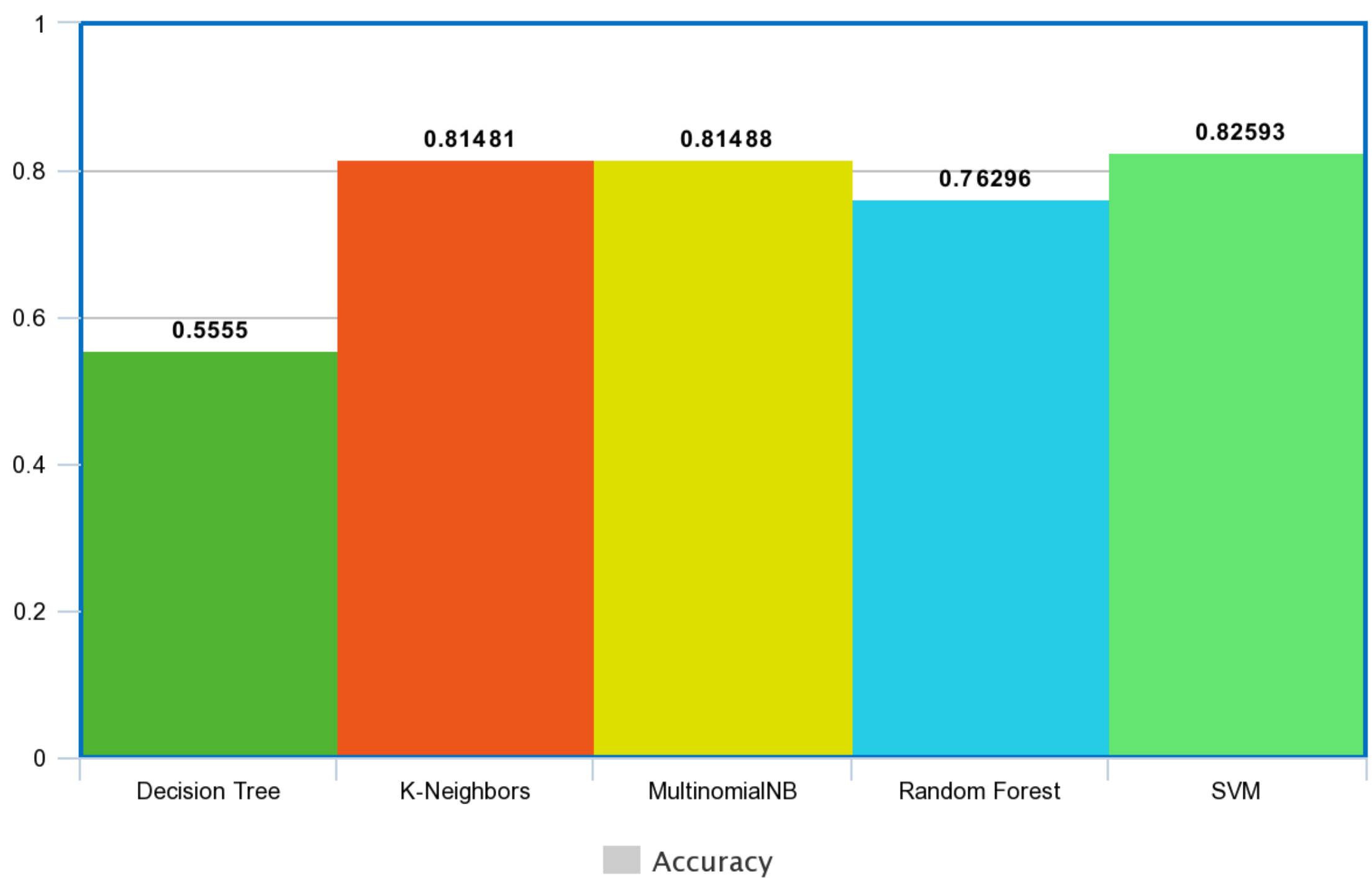


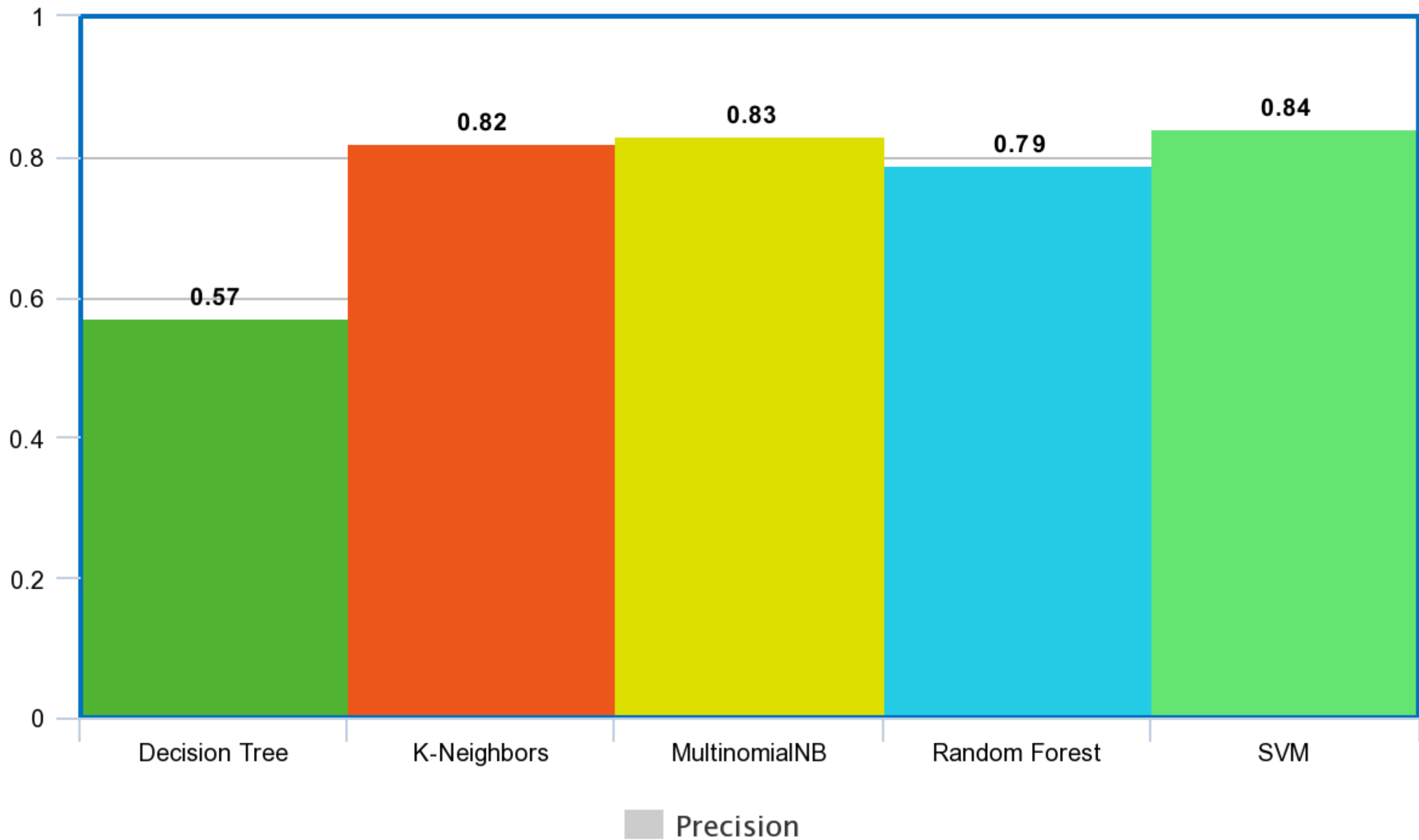
Tested classifiers

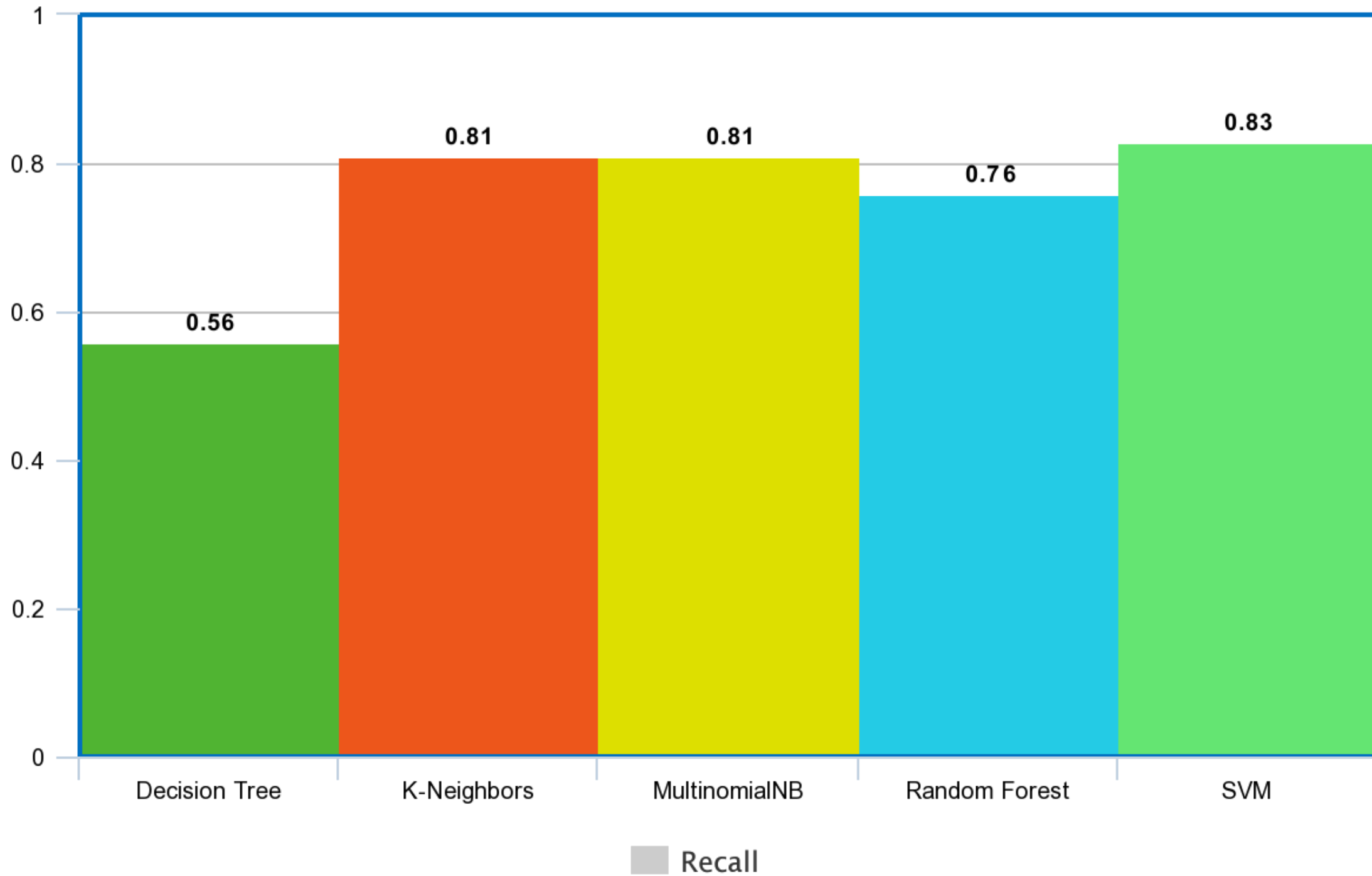
- Decision tree
- KNeighbors
- Multinomial Naive Bayes
- Random Forest
- Support Vector Machines

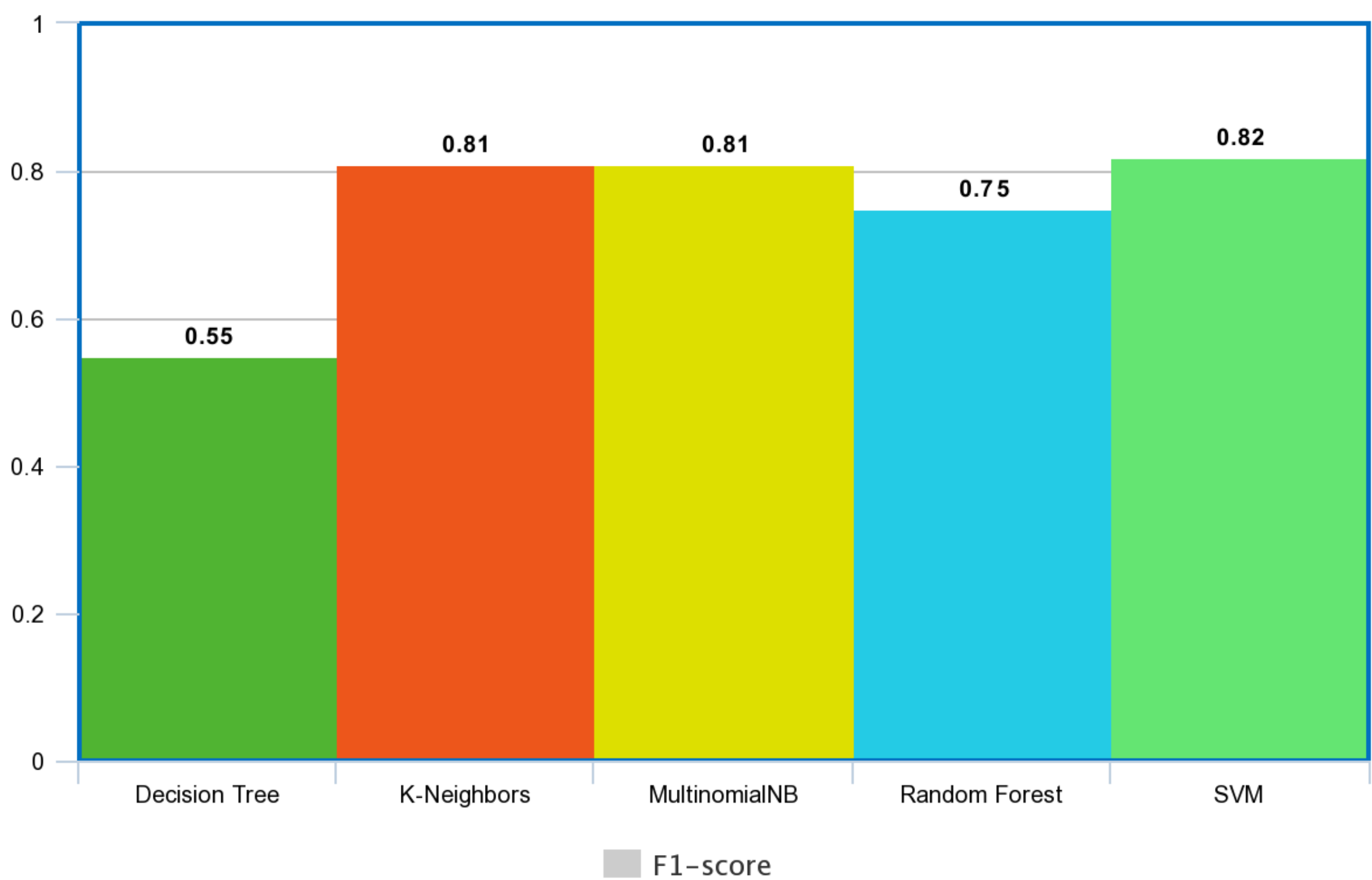
A blue ribbon graphic with a folded end on the left side, containing the word "Results" in white text.

Results









Decision Tree confusion matrix

	am	<u>ac</u>	c	<u>ec</u>	<u>m</u>	p	<u>sc</u>	<u>sp</u>	t
am	[46	2	0	3	4	0	1	2	2]
<u>ac</u>	[4	24	2	4	12	6	2	5	1]
c	[1	2	21	3	8	1	4	14	6]
<u>ec</u>	[3	1	0	38	6	3	2	5	2]
<u>m</u>	[1	4	6	6	28	3	3	5	4]
p	[1	0	1	6	7	40	1	2	2]
<u>sc</u>	[8	5	6	5	8	0	21	4	3]
<u>sp</u>	[0	1	0	4	8	0	1	44	2]
t	[3	0	4	3	4	0	2	6	38]

KNeighbors confusion matrix

	am	<u>ac</u>	c	<u>ec</u>	<u>m</u>	p	<u>sc</u>	<u>sp</u>	t
am	[54	0	0	4	0	1	0	0	1]
<u>ac</u>	[2	38	0	3	11	4	0	1	1]
c	[1	1	43	0	1	2	5	3	4]
<u>ec</u>	[0	4	0	52	1	2	0	0	1]
<u>m</u>	[1	1	4	1	44	7	0	1	1]
p	[0	1	0	0	0	59	0	0	0]
<u>sc</u>	[5	3	1	2	1	1	46	1	0]
<u>sp</u>	[0	2	1	2	0	0	0	55	0]
t	[1	0	0	6	1	2	0	1	49]

MultinomialNB confusion matrix

	am	<u>ac</u>	c	<u>ec</u>	<u>m</u>	p	<u>sc</u>	<u>sp</u>	t
am	[55	0	0	4	0	1	0	0	0]
<u>ac</u>	[1	40	0	3	4	5	3	1	3]
c	[3	4	34	1	0	3	2	2	11]
<u>ec</u>	[0	1	0	55	0	3	0	0	1]
<u>m</u>	[1	7	1	1	40	7	1	1	1]
p	[0	0	0	0	0	60	0	0	0]
<u>sc</u>	[5	3	1	1	0	1	49	0	0]
<u>sp</u>	[0	0	0	5	0	0	0	55	0]
t	[0	0	1	3	1	2	1	0	52]

Random Forest confusion matrix

	am	<u>ac</u>	c	<u>ec</u>	<u>m</u>	p	<u>sc</u>	<u>sp</u>	t
am	[55	0	0	2	0	1	1	0	1]
<u>ac</u>	[0	35	0	7	5	6	2	2	3]
c	[2	2	24	0	8	2	1	11	10]
<u>ec</u>	[0	1	0	50	2	2	0	3	2]
<u>m</u>	[1	4	1	1	42	4	0	4	3]
p	[0	0	0	0	2	58	0	0	0]
<u>sc</u>	[8	4	0	0	3	1	35	5	4]
<u>sp</u>	[0	0	0	0	0	0	0	60	0]
t	[2	0	0	1	2	0	0	2	53]

SVM confusion matrix

	am	<u>ac</u>	c	<u>ec</u>	<u>m</u>	p	<u>sc</u>	<u>sp</u>	t
am	[57	0	0	2	0	1	0	0	0]
<u>ac</u>	[1	42	1	3	8	2	0	1	2]
c	[3	6	30	1	9	0	0	4	7]
<u>ec</u>	[0	1	0	53	0	1	0	0	5]
<u>m</u>	[1	2	0	0	54	1	0	1	1]
p	[0	0	1	1	1	57	0	0	0]
<u>sc</u>	[8	4	0	0	4	1	40	2	1]
<u>sp</u>	[0	0	0	1	0	0	0	59	0]
t	[0	0	0	1	4	0	0	1	54]

A blue ribbon graphic with a 3D effect, featuring a dark blue shadow on the left side. The word "Conclusions" is written in white text on the front of the ribbon.

Conclusions

Conclusions

- We can achieve great performances choosing the right classifier (17.407% of error using the SVM classifier).
- The performance achieved using algorithms based on binary trees are much lower than the others.
- The model is static, and the performances of the classifiers could decrease over time