Graduation Project

Data Mining: Applying and comparing the performance of classification algorithms on a clinical dataset.

Supervised By:

Dr. Ahmad Shraideh

Presented By the Corona Graduates :

Abrar Mohammad

Dania Odeh

The Idea of the Project

The aim of our project Is to analyze and apply prediction algorithms on a clinical dataset obtained from Razan Center for Infertility.

OUR JOURNEY

1

Razan center for infertility

Razan Center for infertility is committed to providing state of the art integrated treatments designed to achieve optimal pregnancy rates. Their aim is to work in partnership with their patients to realize their dreams of parenthood.

Razan Center for Infertility

I.V.F

OUR JOURNEY

2

Our Dataset

We had 10,000 patients' records, and 48 attribute sets. These 10,000 records have gone under the processing step to settle into 4885 patients' records, 51 attribute sets, and one label column that indicates the success or failure of pregnancy after the ICSI treatment.

ICSI Dataset (N:numerical, C: categorical, S: string, B: binary)		
pk(N)	Male factor (B)	HSG N LT(B)
outcome(B)	Gender selection(B)	HSG N RT(B)
M Age(N)	Ovarian unsufficiency (B)	HSG Tubes Hydrosalpinx LT(B)
F Age(N)	PCOS(B)	HSG Tubes Hydrosalpinx RT(B)
Infertility Type(C)	PGD-PCR(B)	Laproscopy Tube LT status(B)
F Blood Type(C)	PGD-FISH(B)	Laproscopy Tube RT status(b)
M Blood Type (C)	Tubal(B)	Laproscopy Ovary LT status(B)
F weight(N)	Unexplained (B)	Laproscopy Ovary RT status(B)
Kinship(C)	F surgical History (S)	Semen Volume (N)
F Height (N)	F current medications(S)	Semen Count(N)
F BMI(N)	M surgical history (S)	Semen Motility (N)
Fcycle status(C)	AFC LT(N)	Semen Morphology(N)
Anovoluation(B)	AFC RT(N)	TESE +ve(B)
Endomitriosis(B)	Uterus status(B)	TESE -ve(B)
Hypo Female (B)	Hystroscopy Cavity status(B)	
Hypo Male (B)	HSG Cavity status(B)	

OUR JOURNEY

3
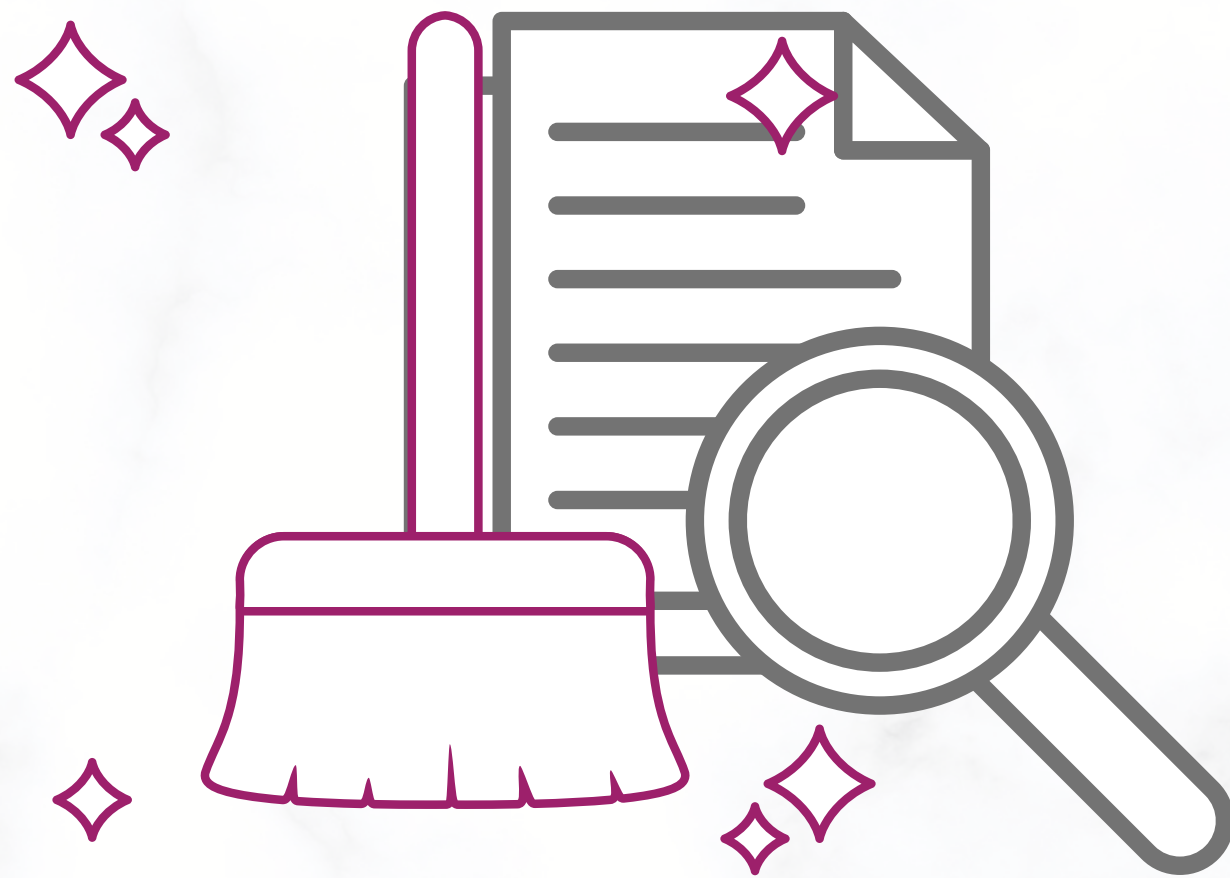
Data Preparation

The data preparation process contains 4 main process to be applied on the various attributes we're handling: replace, transform, filter and categorize.

OUR JOURNEY

4

Modeling

We have applied 4 classification models which are:
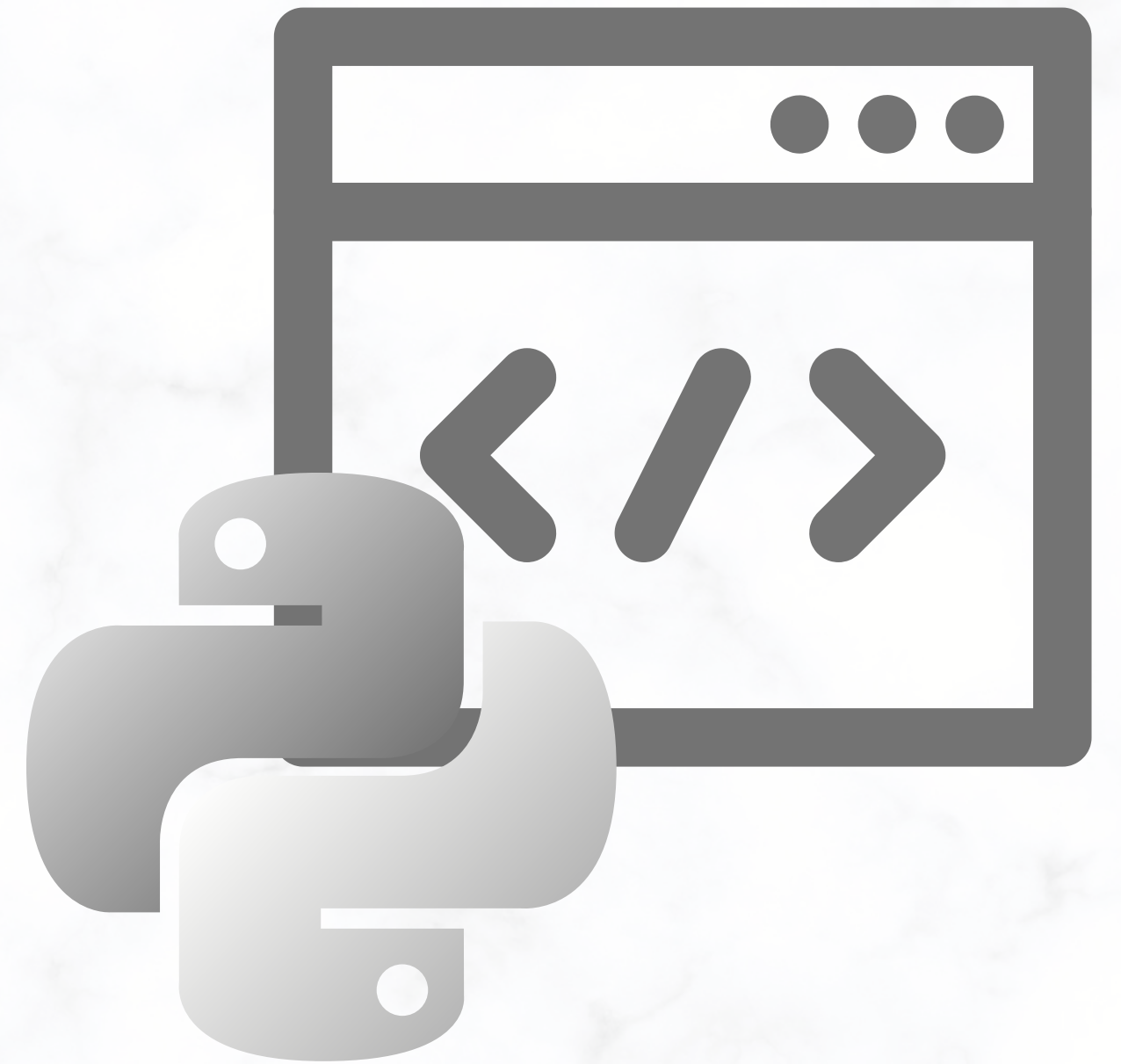
Naive Bayes

Random Forest

K-Nearest Neighbor

Logistic Regression

OUR JOURNEY

5

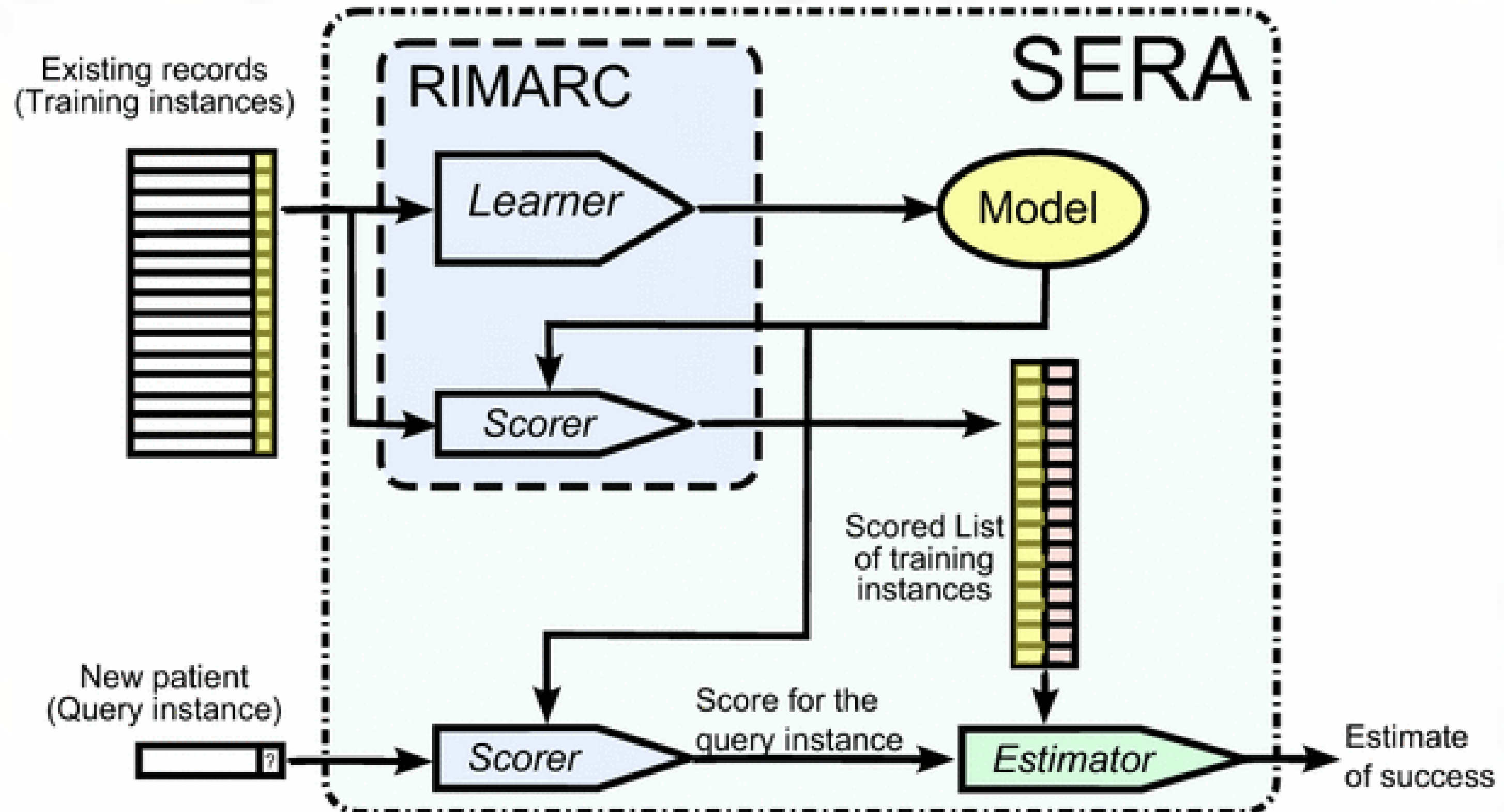# The RIMARC Algorithm

Results Comparison:

algorithm	AUC	Execution time
Naieve Bayes Classifier	0.876	1 second
Random Forest	0.93	2 seconds
k-Nearest Neighbor	0.918	1 second
Logestic Regression	0.93	1 second
RIMARC	0.921	956 ms

OUR JOURNEY

6

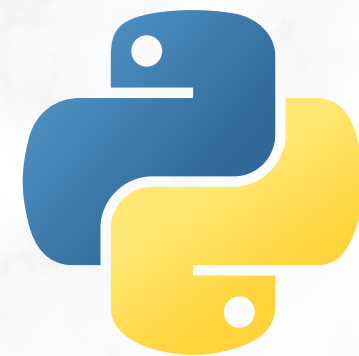# Success Estimation using Ranking Algorithms (SERA)

OUR JOURNEY

7

Research Paper

Tools Used

Thank you!