

# STROKE PREDICTION VIA SUPERVISED METHODS

ARNAU ABELLA AND ANTONI CASAS UNIVERSITAT POLITÈCNICA DE CATALUNYA



## INTRODUCTION

Strokes are an unexpected killer and hard to predict, being the second leading cause of death globally. Therefore, the detection of segments of the population at risk could enhance the ability of hospitals and other medical teams to offer early treatments and screenings to reduce this risk.

Our goal is to predict whether or not an individual is going to suffer a stroke.

## MATERIALS & METHODS

The stroke prediction dataset is composed by 11 features: 3 numerical and 8 categorical. The dataset only has 5110 observations. The binary classification problem is highly imbalanced, with only a 5% of the population belonging to the positive class.

To deal with the imbalance, we applied *SMOTE with Tomek Links* which applies the over-sampling method SMOTE combined with the under-sampling method Tomek links. The combination of both should re-balance the dataset while avoiding the creation of noisy samples from the interpolation of outliers and inliers.

We have also explored factor analysis, in particular MCA, and dimensionality.

We considered the following models:

- Logistic Regression with Regularization
- Bayesian Models
- Support Vector Machines
- Decision Trees
- Ensemble Methods: Random Forest, Ada Boosting, and Voting Classifiers
- Neural Networks

We focused on achieving the best *F1-score* while keeping the *recall* high.

## REFERENCES

- [1] Paytonfisher. Stroke prediction ml project - 96% score, Jun 2021.
- [2] Mariahchristy. Stroke dataset: Eda and prediction: 96% accuracy, Jun 2021.

## RESULTS 2

Donec faucibus purus at tortor egestas eu fermentum dolor facilisis. Maecenas tempor dui eu neque fringilla rutrum. Mauris *lobortis* nisl accumsan.

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 1: Table caption

Nulla ut porttitor enim. Suspendisse venenatis dui eget eros gravida tempor. Mauris feugiat elit et augue placerat ultrices. Morbi accumsan enim nec tortor consectetur non commodo.

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 2: Table caption

## FUTURE RESEARCH

Integer sed lectus vel mauris euismod suscipit. Praesent a est a est ultricies pellentesque. Donec tincidunt, nunc in feugiat varius, lectus lectus auctor lorem, egestas molestie risus erat ut nibh.

## RESULTS 1

Result's Table						
MLP(default)	75.2%	73.2%	19.2%	85.4%	52.3%	74.3%
LR(default)	74.6%	73.2%	18.8%	84.9%	51.8%	73.9%
SVC(default)	80.8%	43.9%	15.5%	89.2%	52.4%	63.1%
RF(default)	75.3%	51.2%	14.3%	85.6%	49.9%	63.8%
DT(default)	75.0%	46.3%	13.0%	85.4%	49.2%	61.3%
Knn(default)	81.4%	34.1%	12.8%	89.6%	51.2%	58.8%
AdaBoost(default)	42.9%	90.2%	11.2%	57.9%	34.6%	65.6%
GB(default)	34.4%	95.1%	10.4%	48.3%	29.4%	63.5%
QDA(default)	96.0%	0.0%	0.0%	98.0%	49.0%	50.0%
NB(default)	96.0%	0.0%	0.0%	98.0%	49.0%	50.0%
	Accuracy	Recall	F1-score (class 1)	F1-score (class 0)	F1-score (macro avg)	ROC AUC Score

Figure 1: Figure caption

Sed fringilla tempus hendrerit. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Etiam ut elit sit amet metus lobortis consequat sit amet in libero. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus vel sem magna. Nunc at convallis urna. isus ante. Pellentesque condimentum dui. Etiam sagittis purus non tellus tempor volutpat. Donec et dui non massa tristique adipiscing. Quisque vestibulum eros eu.

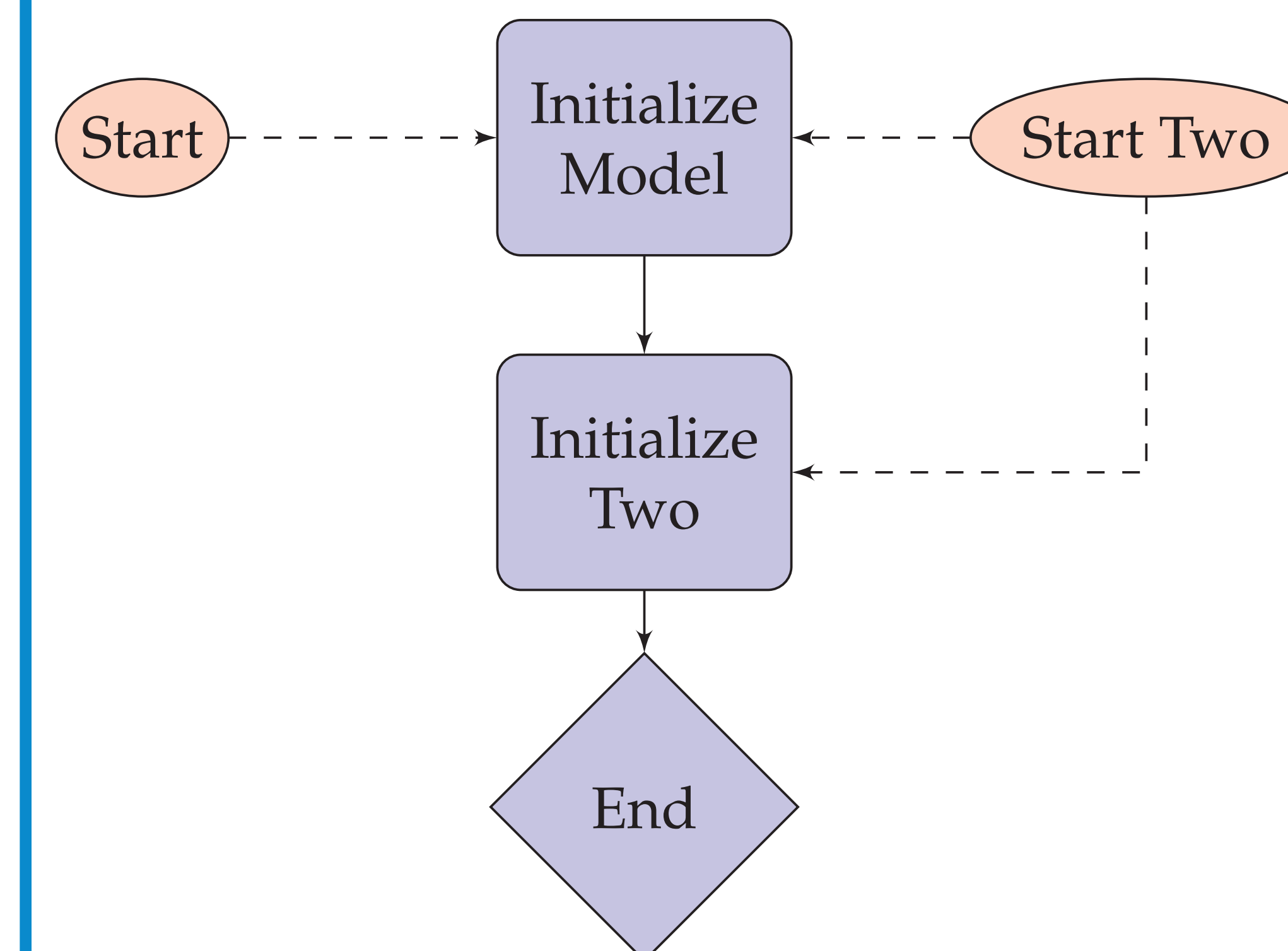
Aliquam auctor, metus id ultrices porta, risus enim cursus sapien, quis iaculis sapien tortor sed odio. Mauris ante orci, euismod vitae tincidunt eu, porta ut neque. Aenean sapien est, viverra vel lacinia nec, venenatis eu nulla. Maecenas ut nunc nibh, et tempus libero. Aenean vitae risus ante. Pellentesque condimentum dui. Etiam sagittis purus non tellus tempor volutpat. Donec et dui non massa tristique adipiscing.

lum eros eu.

Result's Table						
MLP(default)	75.2%	73.2%	19.2%	85.4%	52.3%	74.3%
LR(default)	74.6%	73.2%	18.8%	84.9%	51.8%	73.9%
SVC(default)	80.8%	43.9%	15.5%	89.2%	52.4%	63.1%
RF(default)	75.3%	51.2%	14.3%	85.6%	49.9%	63.8%
DT(default)	75.0%	46.3%	13.0%	85.4%	49.2%	61.3%
Knn(default)	81.4%	34.1%	12.8%	89.6%	51.2%	58.8%
AdaBoost(default)	42.9%	90.2%	11.2%	57.9%	34.6%	65.6%
GB(default)	34.4%	95.1%	10.4%	48.3%	29.4%	63.5%
QDA(default)	96.0%	0.0%	0.0%	98.0%	49.0%	50.0%
NB(default)	96.0%	0.0%	0.0%	98.0%	49.0%	50.0%
	Accuracy	Recall	F1-score (class 1)	F1-score (class 0)	F1-score (macro avg)	ROC AUC Score

Figure 2: Figure caption

## CONCLUSION



- Pellentesque eget orci eros. Fusce ultricies, tellus et pellentesque fringilla, ante massa luctus libero, quis tristique purus urna nec nibh. Phasellus fermentum rutrum elementum. Nam quis justo lectus.
- Vestibulum sem ante, hendrerit a gravida ac, blandit quis magna.
- Donec sem metus, facilisis at condimentum eget, vehicula ut massa. Morbi consequat, diam sed convallis tincidunt, arcu nunc.
- Nunc at convallis urna. isus ante. Pellentesque condimentum dui. Etiam sagittis purus non tellus tempor volutpat. Donec et dui non massa tristique adipiscing.

## CONTACT INFORMATION

**Web** [www.university.edu/smithlab](http://www.university.edu/smithlab)

**Email** [john@smith.com](mailto:john@smith.com)

**Phone** +1 (000) 111 1111