

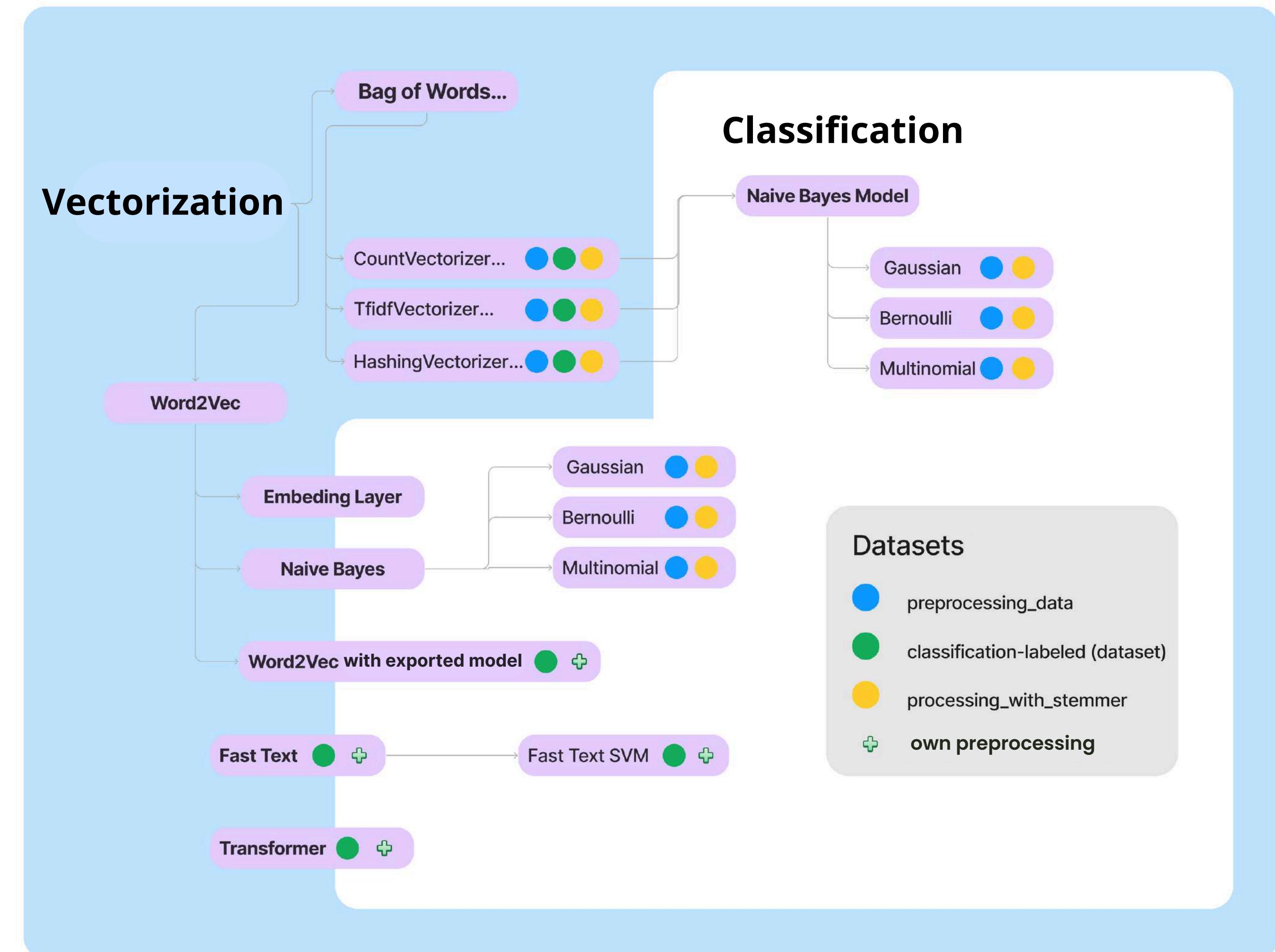




Jäger

Sprint 4 - Construction of the API with a highlighted model

Models overview





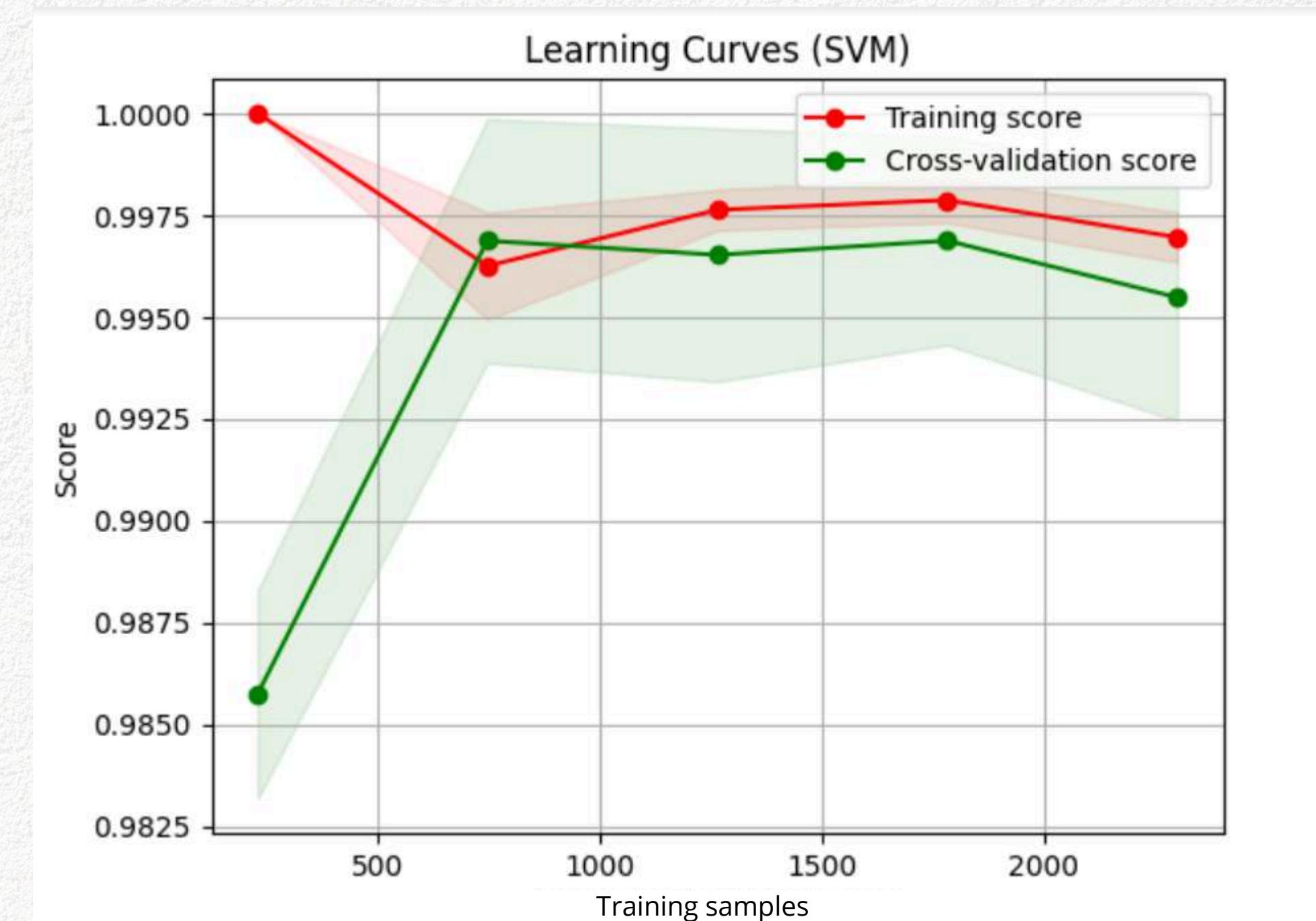
Sprint 3 Review

FastText

Accuracy: 57,17%
Precision: 75,81%
Recall: 57,17%
F1-Score: 58,08%

FastText + SVM

Accuracy: 98,78%
Precision: 98,79%
Recall: 98,78%
F1-Score: 98,78%





Boot

- A 3x4 grid of 12 blue dots arranged in three rows and four columns.

Sprint 4

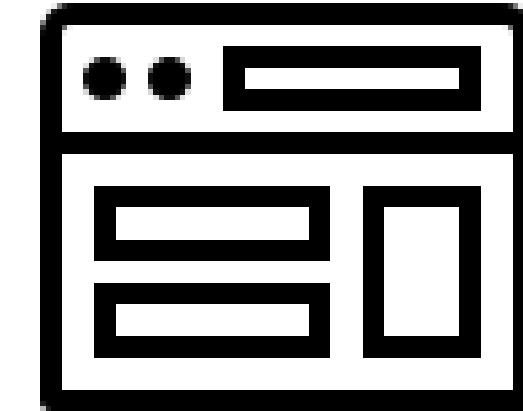
Model Enhancement



API development



User Experience



Boot

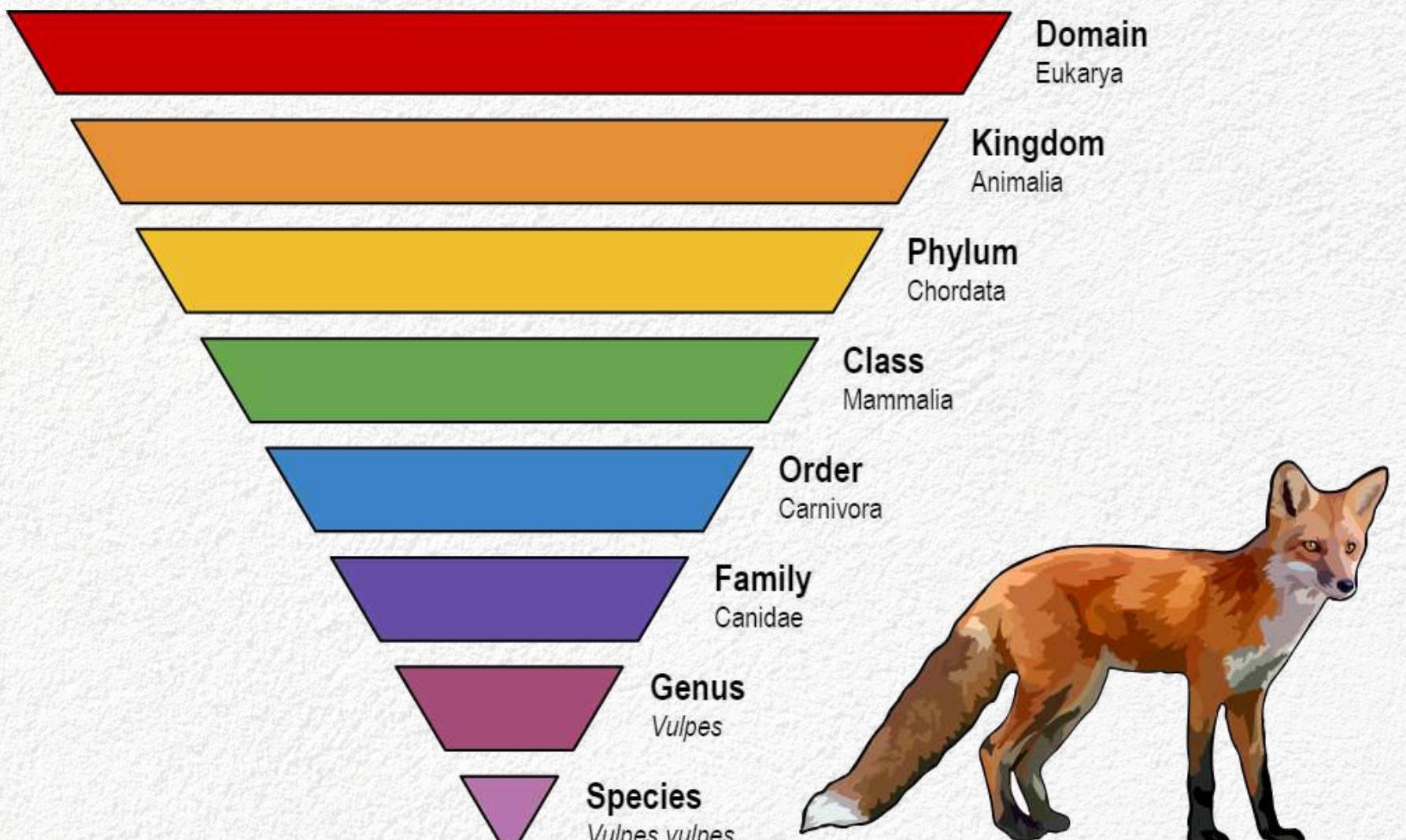


Solution

Fast Text + Support Vector Machine

FastText

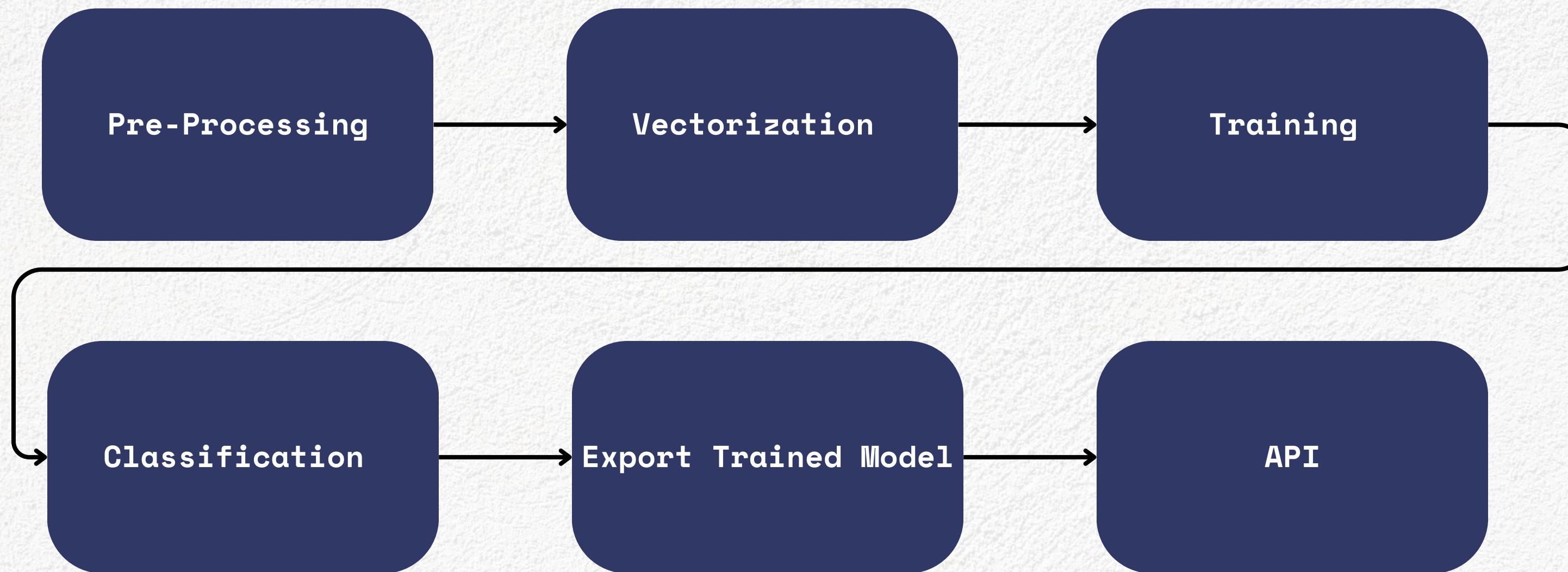
- Facebook AI Research (FAIR)
- Considers words as a **combination of subwords**
- **Bigrams** were made to identify compound words
- **Soft Max Hierarchical** in which it divide words as if it were a taxonomy tree
- **Quick to train**



Red fox (*Vulpes vulpes*)

Solution

Fast Text + SVM



Boot

Solution

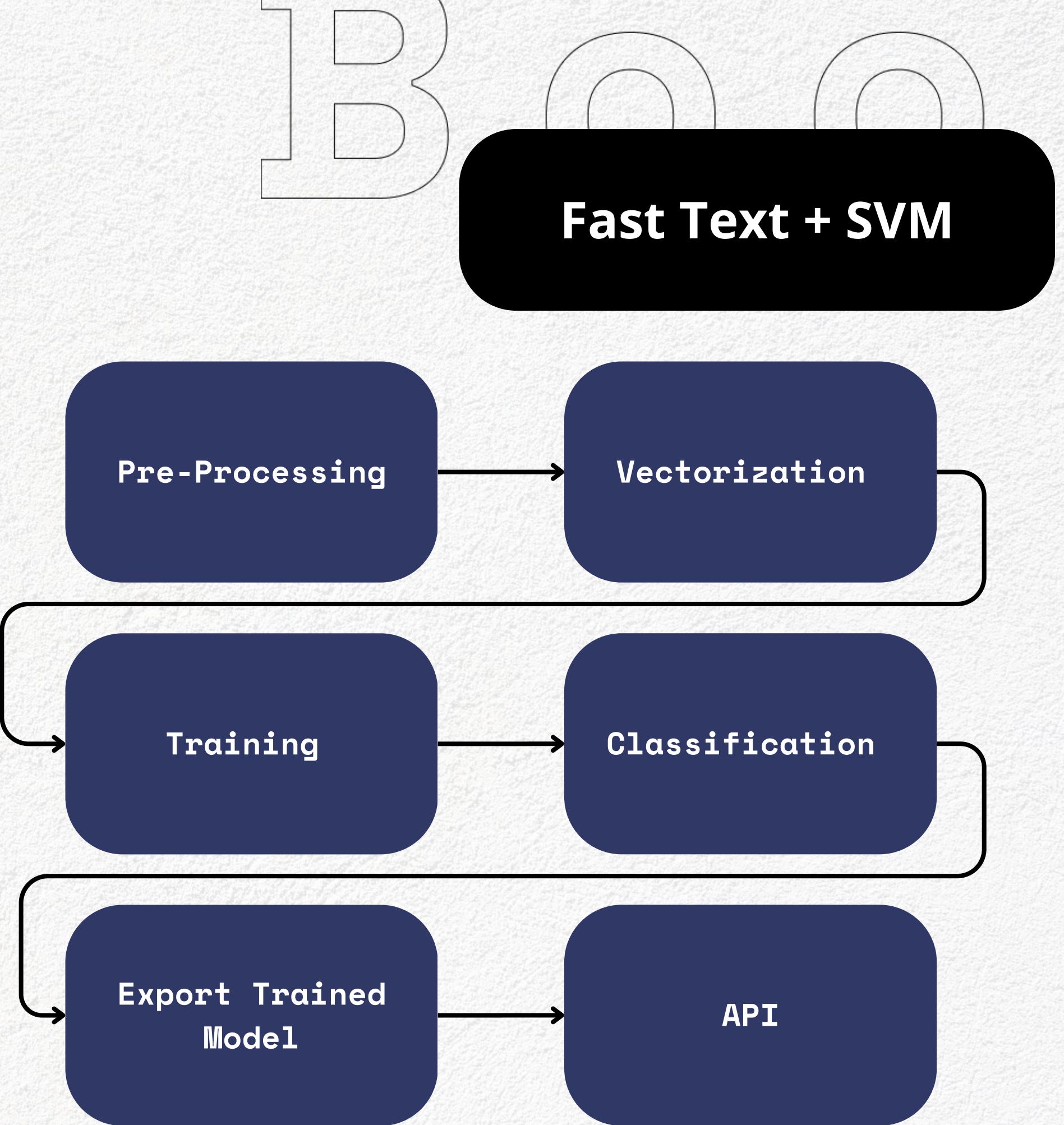
Fast Text + Support Vector Machine

SVM

- The SVM classifies, being a **dimensional classifier**
- Creates a hyperplane, a type of map divides the plane into two parts.

Kernels:

- Calculates the **similarity** between two points
- RBF type (**Radial Basis Function**): for non-linearly separable data in a higher dimensional space

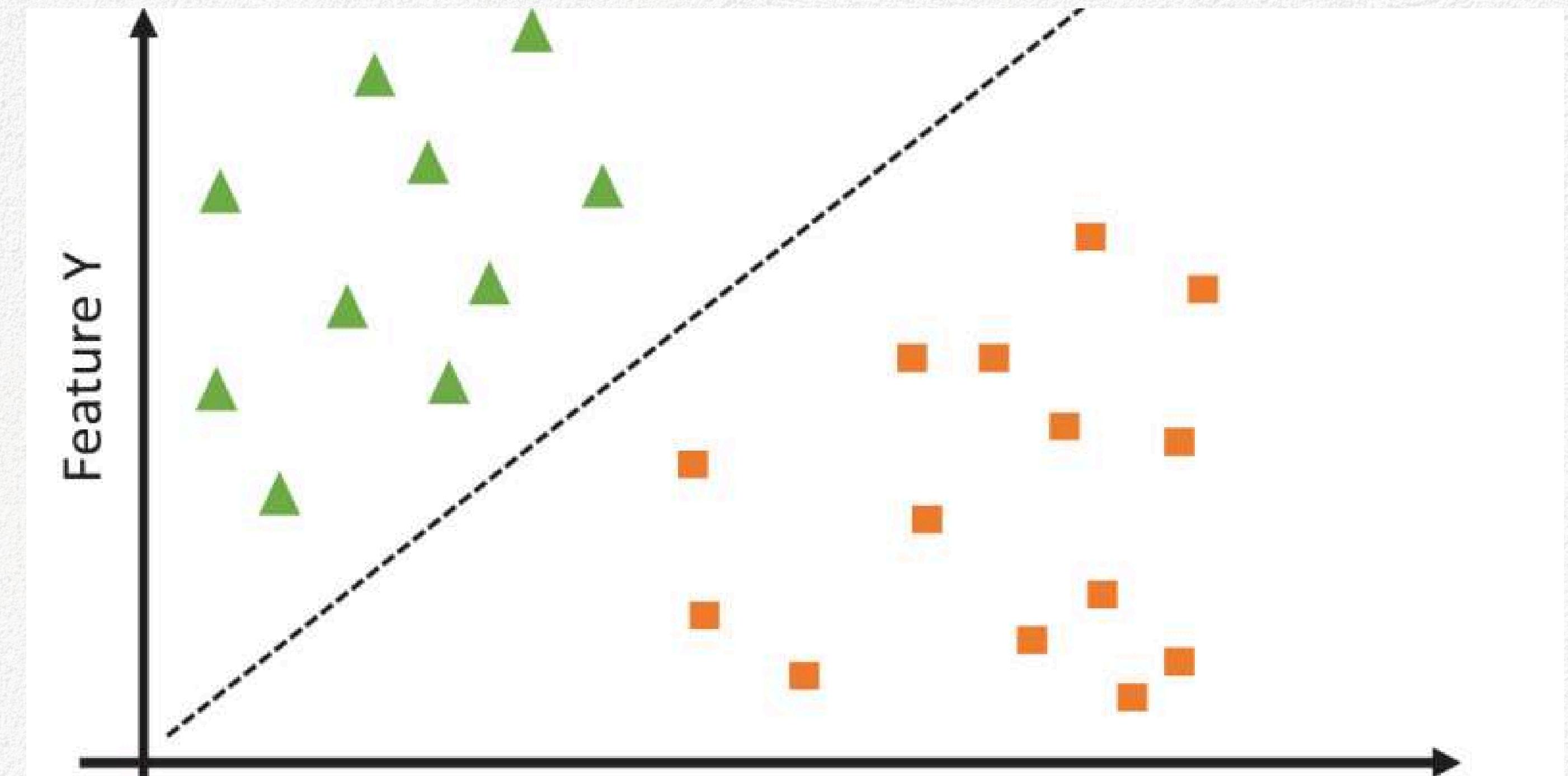


Fast Text + SVM

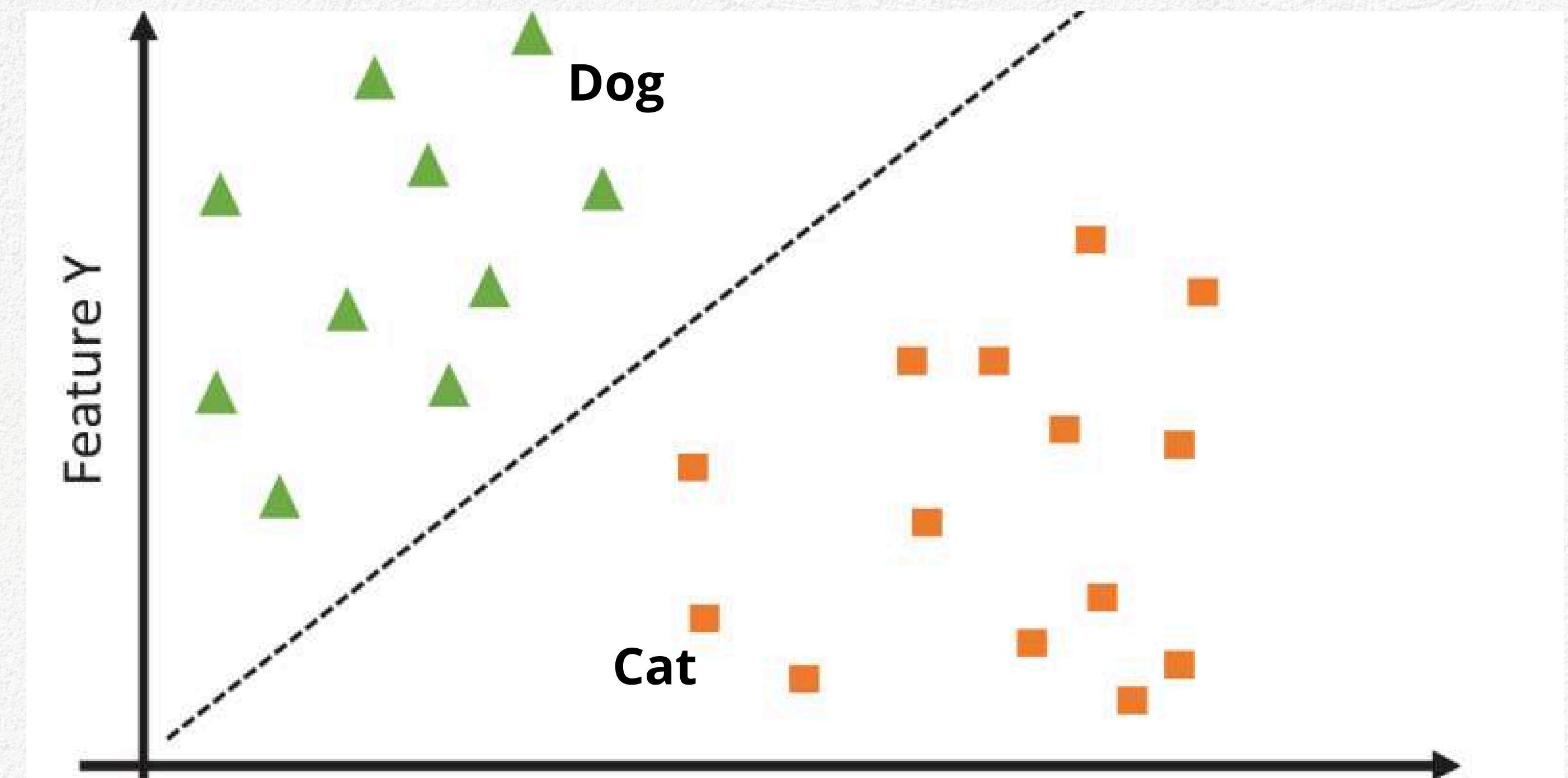
Boot

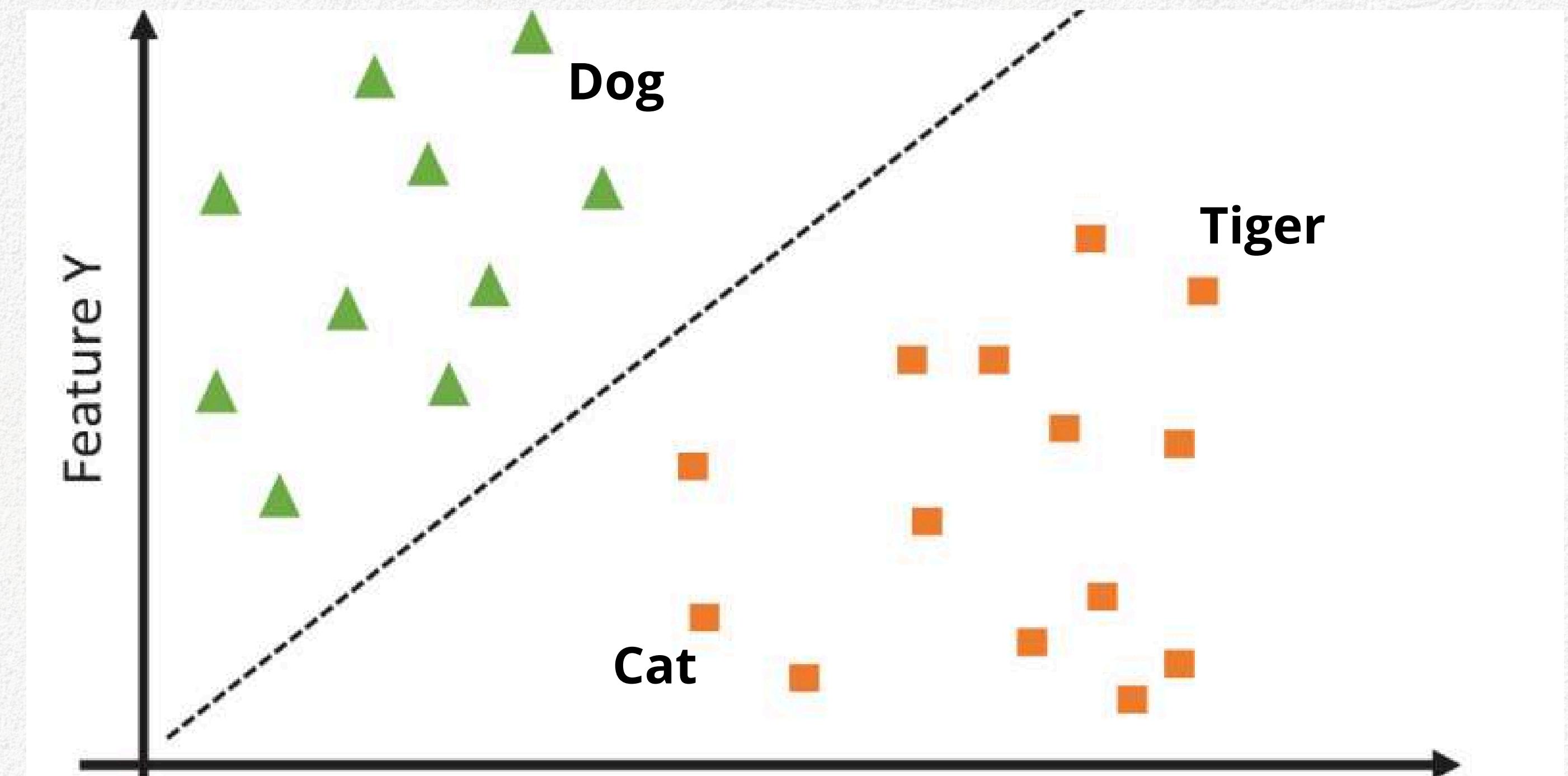


Boot



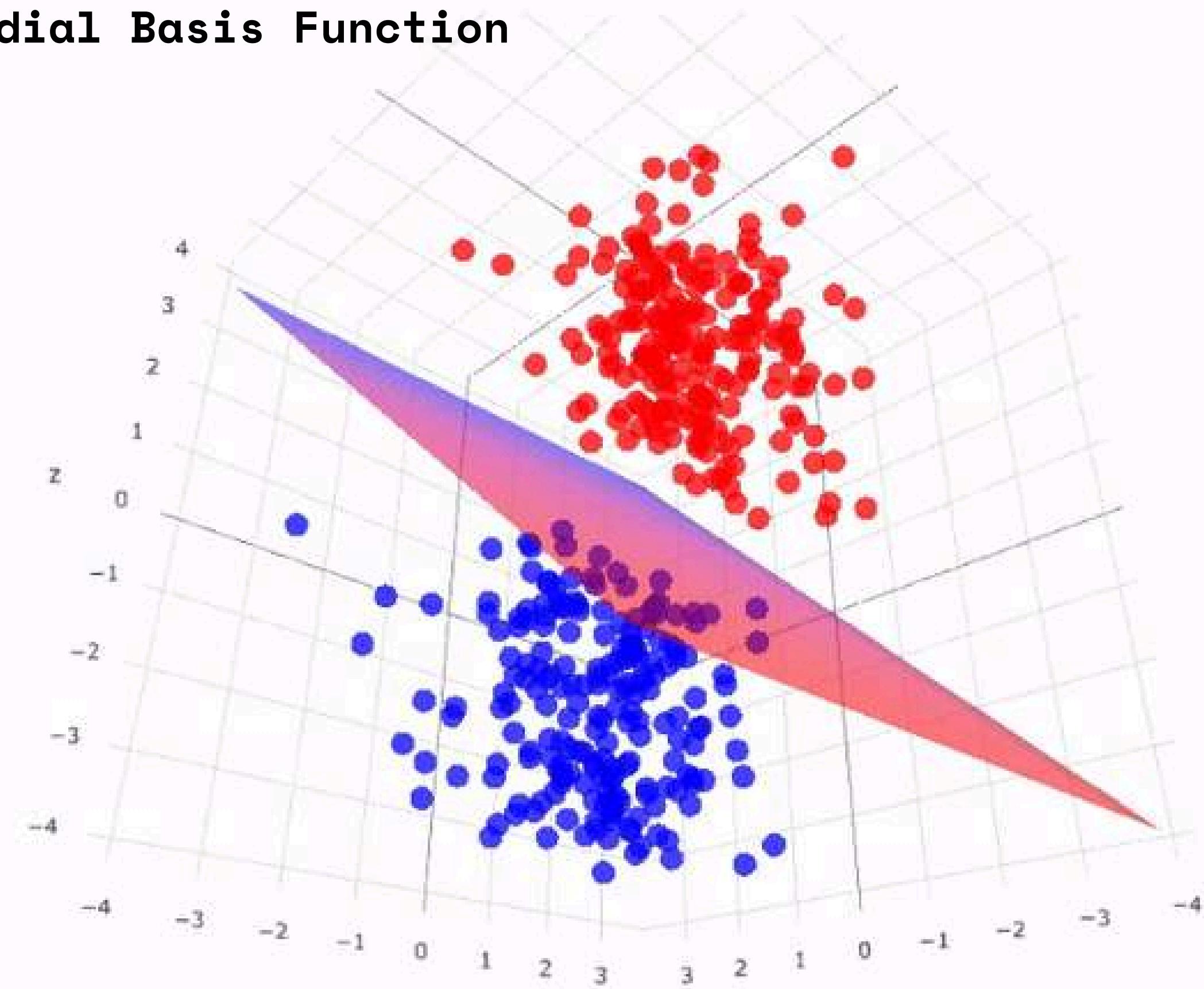
Boot







Radial Basis Function



New results

Accuracy: 93.06%

Precision: 93.32%

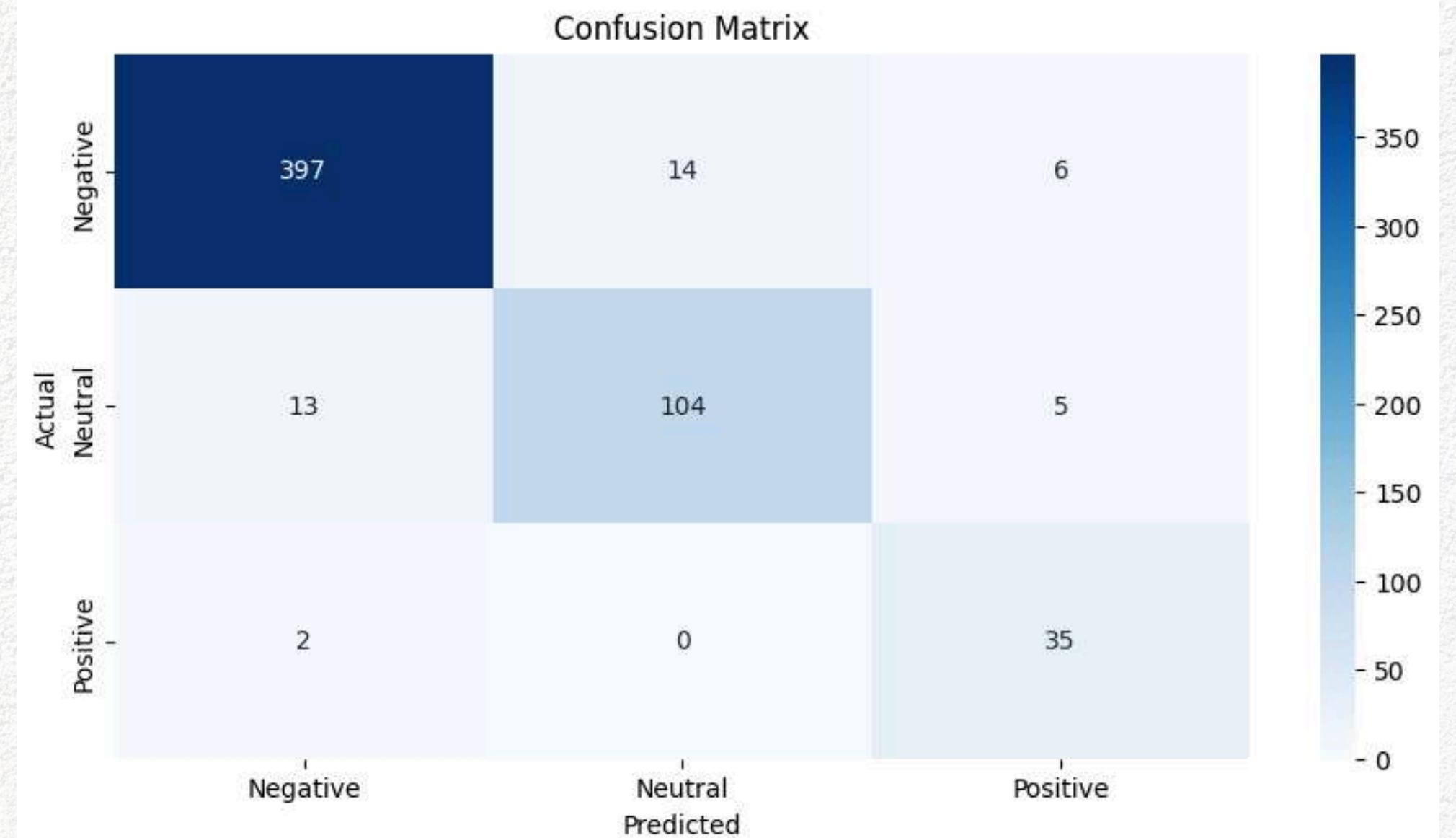
Recall: 93.06%

F1-score: 93.06%

Cohen Kappa: 83.99%

Youden's Index: 86.83%

Fast Text + SVM



Boot



Why the model ?

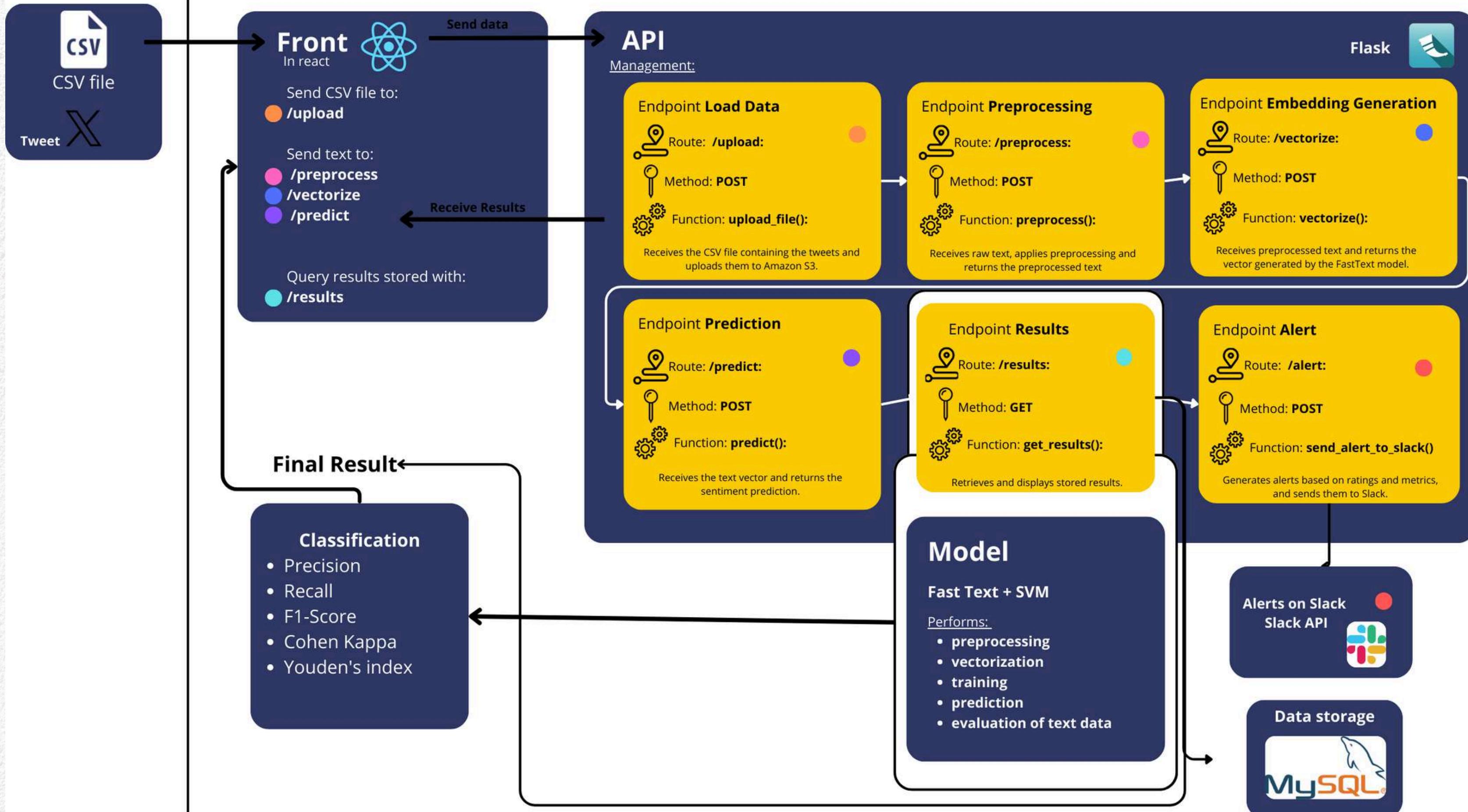
Disadvantages :

- Loss of Sentence Syntax (FastText)
- Computational Complexity (SVM)
- Sensitivity to Unbalanced Data (SVM)
- **Model explainability (SVM)**

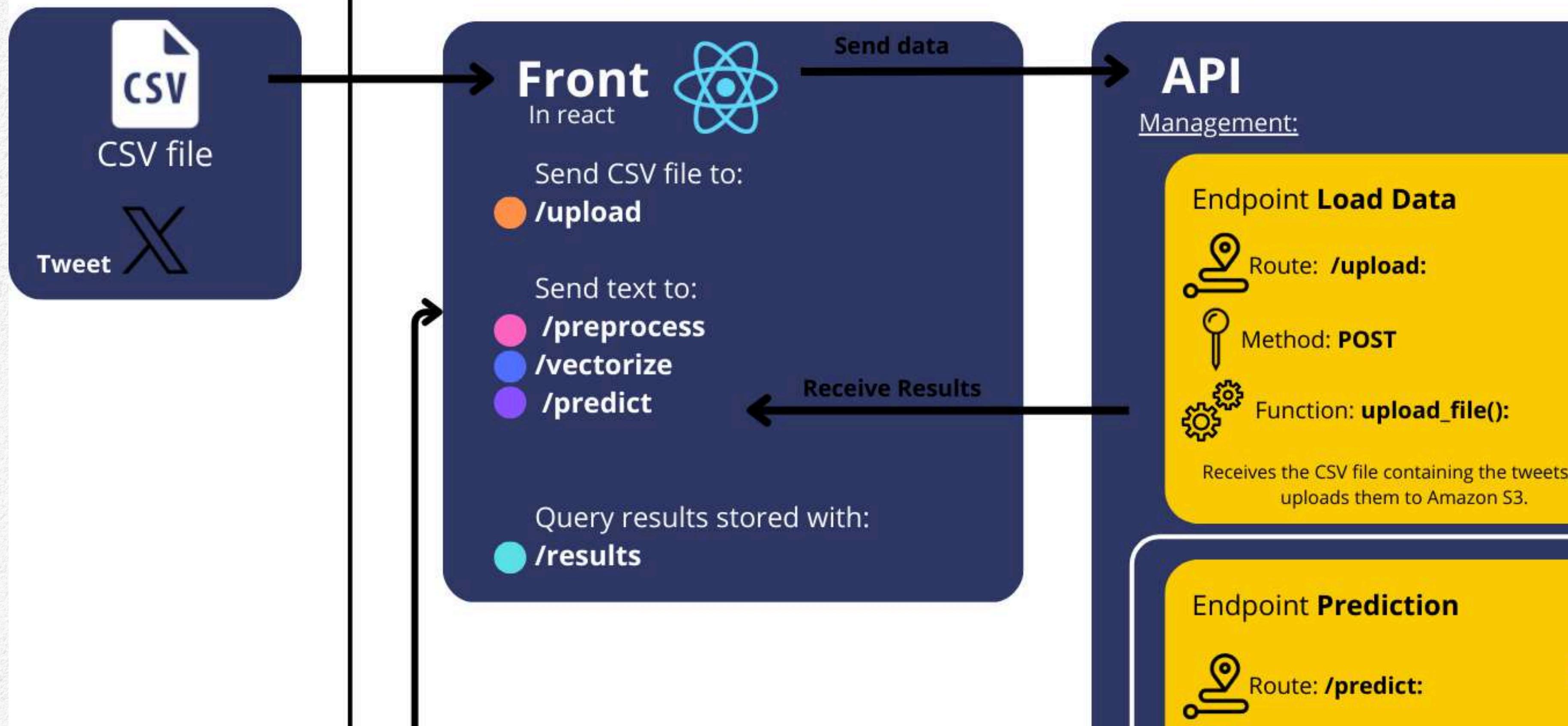
Benefits :

- Computational Efficiency (FastText)
- Fast training with low complexity (FastText)
- **Rich Semantic Representations** (FastText)
- Training speed (FastText)
- **Data volume** (FastText)
- Generalization Capability(SVM)

Solution Macro Architecture



Solution Macro Architecture



API

Management:



Flask

Endpoint Load Data



Route: /upload:



Method: POST



Function: upload_file():

Receives the CSV file containing the tweets and uploads them to Amazon S3.

Endpoint Preprocessing



Route: /preprocess:



Method: POST



Function: preprocess():

Receives raw text, applies preprocessing and returns the preprocessed text

Endpoint Embedding Generation



Route: /vectorize:



Method: POST



Function: vectorize():

Receives preprocessed text and returns the vector generated by the FastText model.

Endpoint Prediction



Route: /predict:



Method: POST



Function: predict():

Receives the text vector and returns the sentiment prediction.

Endpoint Results



Route: /results:



Method: GET



Function: get_results():

Retrieves and displays stored results.

Endpoint Alert



Route: /alert:



Method: POST



Function: send_alert_to_slack()

Generates alerts based on ratings and metrics, and sends them to Slack.

Model

Endpoint Prediction

Route: /predict:

Method: POST

Function: `predict()`:

Receives the text vector and returns the sentiment prediction.

Endpoint Results

Route: /results:

Method: GET

Function: `get_results()`:

Retrieves and displays stored results.

Endpoint Alert

Route: /alert:

Method: POST

Function: `send_alert_to_slack()`

Generates alerts based on ratings and metrics, and sends them to Slack.

Model

Fast Text + SVM

Performs:

- preprocessing
- vectorization
- training
- prediction
- evaluation of text data

Alerts on Slack

Slack API



Data storage



Endpoint Prediction



Route: /predict:



Method: POST



Function: predict():

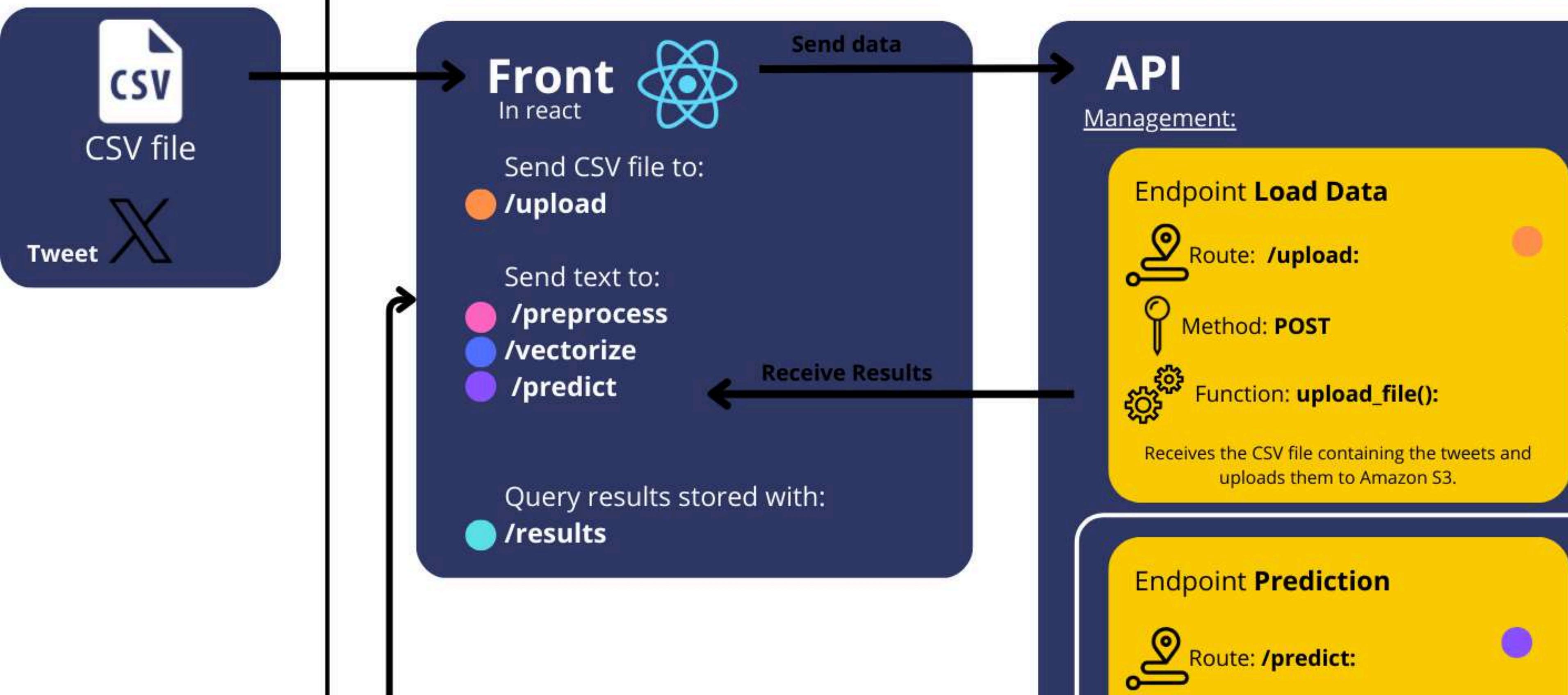
Receives the text vector and returns the sentiment prediction.

Final Result

Classification

- Precision
- Recall
- F1-Score
- Cohen Kappa
- Youden's index

Solution Macro Architecture



API demo



Boot

Simulação

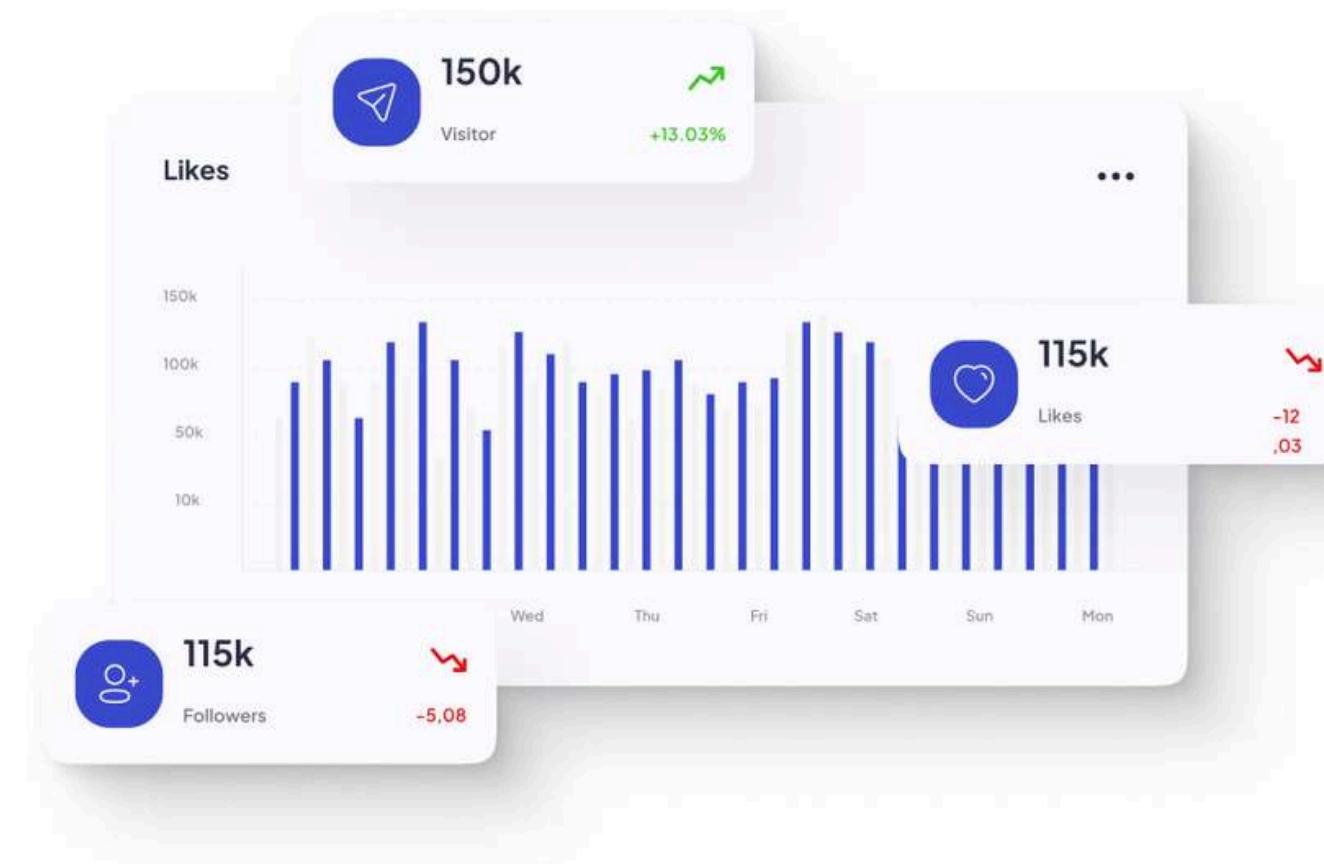
Input

Sobre

Jäger

Jäger, developed by the Boot group, is a simulation tool for sentiment analysis using Natural Language Processing (NLP) techniques. With Jäger, you can evaluate and classify emotions in texts, visualize data intuitively, and generate detailed reports. Discover how our solution can mitigate risks and assist in decision-making through simulation!

Simulate



Boot



Boot sentiment analyst

Jäger

Simply input your text, and our tool will instantly classify it as positive, negative, or neutral. Using advanced Natural Language Processing (NLP), we provide quick and accurate sentiment analysis to help you understand the emotional tone of any message. Try it now and gain valuable insights from your text!

 Input



Boot

Simulação

Input

Sobre

Boot sentiment analyst

Jäger

Simply input your text, and our tool will instantly classify it as positive, negative, or neutral. Using advanced Natural Language Processing (NLP), we provide quick and accurate sentiment analysis to help you understand the emotional tone of any message. Try it now and gain valuable insights from your text!



Intensity of feeling - 07

The tweet is classified as negative !

Boot

**Serena**

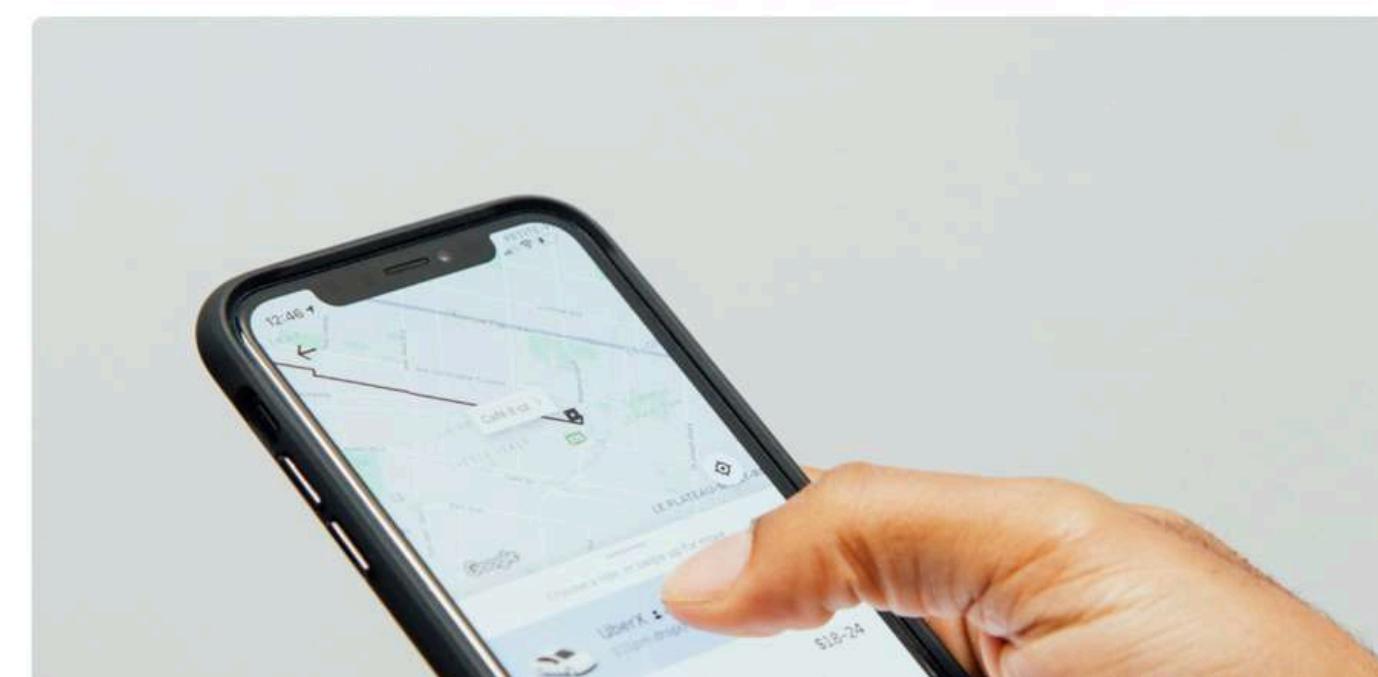
@serena_example

Uber's service has really gone downhill lately. The wait times are longer, the fares are higher, and the drivers seem less professional. It's frustrating to pay more for a service that used to be reliable and efficient.

 177 784 784**Sarah**

@sarah_example

Recently, the quality of Uber's service has significantly declined. Wait times are longer, fares have increased, and drivers seem less professional. It's disappointing to pay more for a service that used to be reliable and efficient.



Recently, the quality of Uber's service has significantly declined. Wait times are longer, fares have increased, and drivers seem less professional. It's disappointing to pay more for a service that used to be reliable and efficient.



177

784

129



Josh

@josh_example

Uber's service has really gone downhill lately. The wait times are longer, the fares are higher, and the drivers seem less professional. It's frustrating to pay more for a service that used to be reliable and efficient.

177

784

784

Boot



Solution

Critic comment alert:

- Level: Brand Attack
- Link: <https://sample.com/>
- Platform: Twitter
- Related service: Uber Eats
- Post type: Media Post

Number of posts with the same classification: 256

Positive trend alert:

- Trend: Service Quality
- Link: <https://sample.com/>
- Platform: Twitter
- Related service: Uber
- Post type: Media Post

Number of posts with the same classification: 256

Improvement comment alert:

- Topic: Product Improvement
- Link: <https://sample.com/>
- Platform: Twitter
- Related service: Uber Eats
- Post type: Media Post

Number of posts with the same classification: 256

Sprint Conclusions

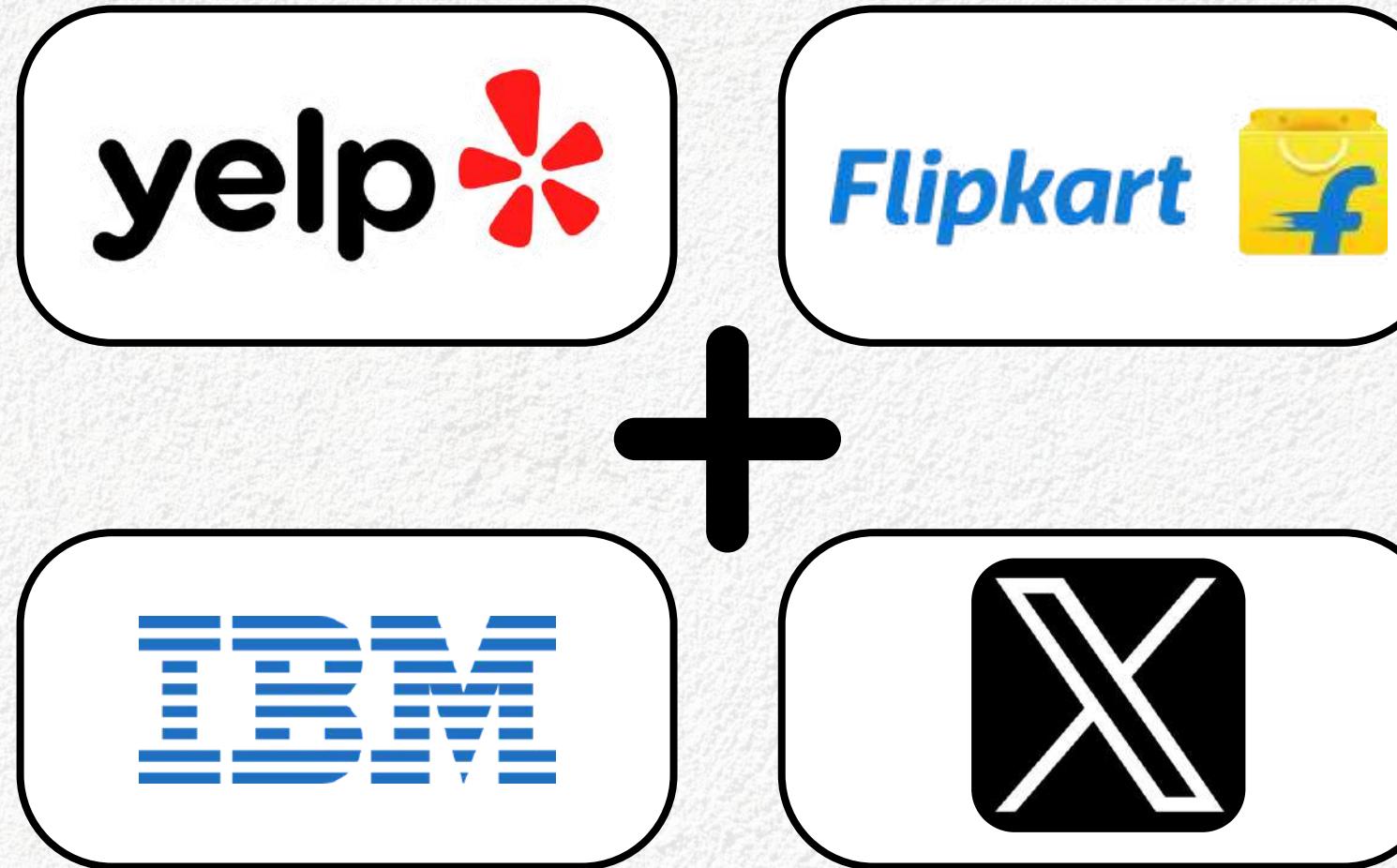
We can still improve a lot in relation to the model
Test other possibilities with the same model



Metrics	Sprint 3	Sprint 4
Accuracy	98,78%	93.06%
Precision	98,78%	93.32%
Recall	98,78%	93.06%
F1-score	98,78%	93.06%

Different approaches

kaggle



200.000 Data

Next Steps : Sprint 5



Deploy the best model



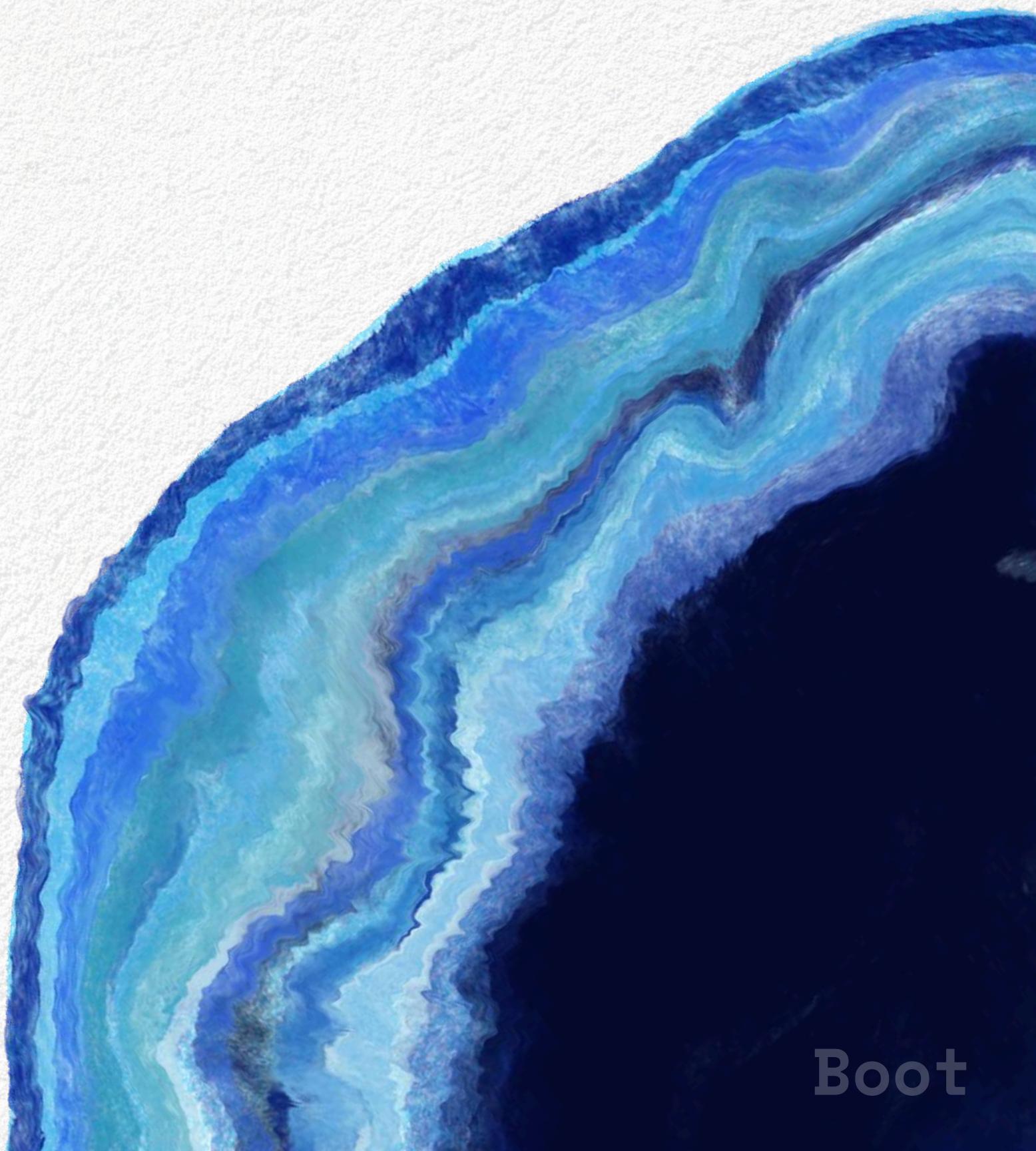
Finish the API



Integrate with Front-end



Slack Full integration





Boot



Matheus R.



Henrique Cox



Eduardo HOS



Mateus M.



Marcelo Saasi



Otto



Celine



Boot