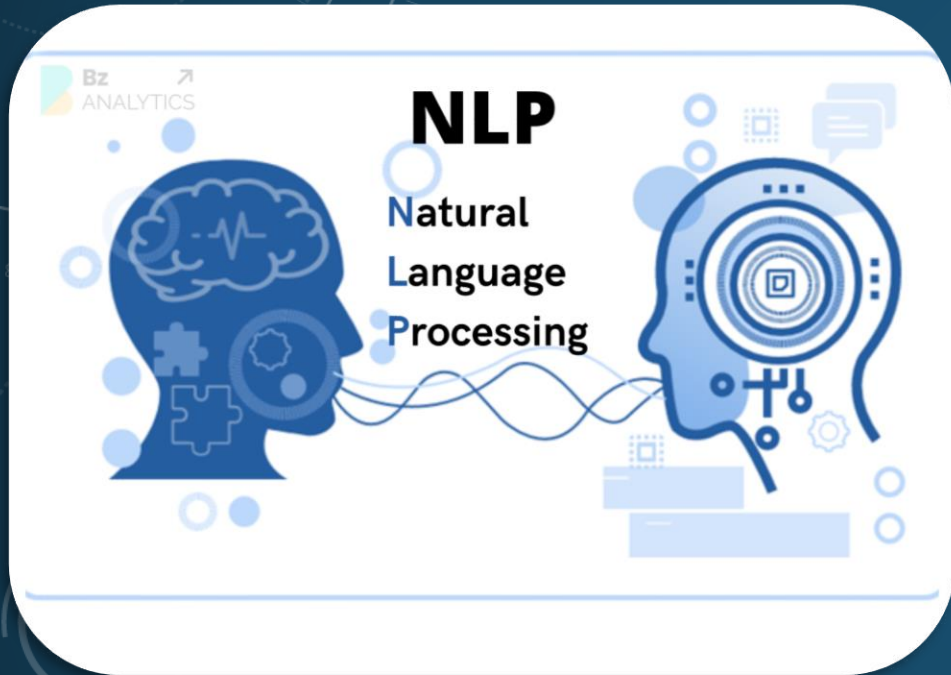# A GUIDE TO RECOGNIZING SPECIFIC PSYCHOLOGICAL LABELS FROM TEXT USING ML & AI
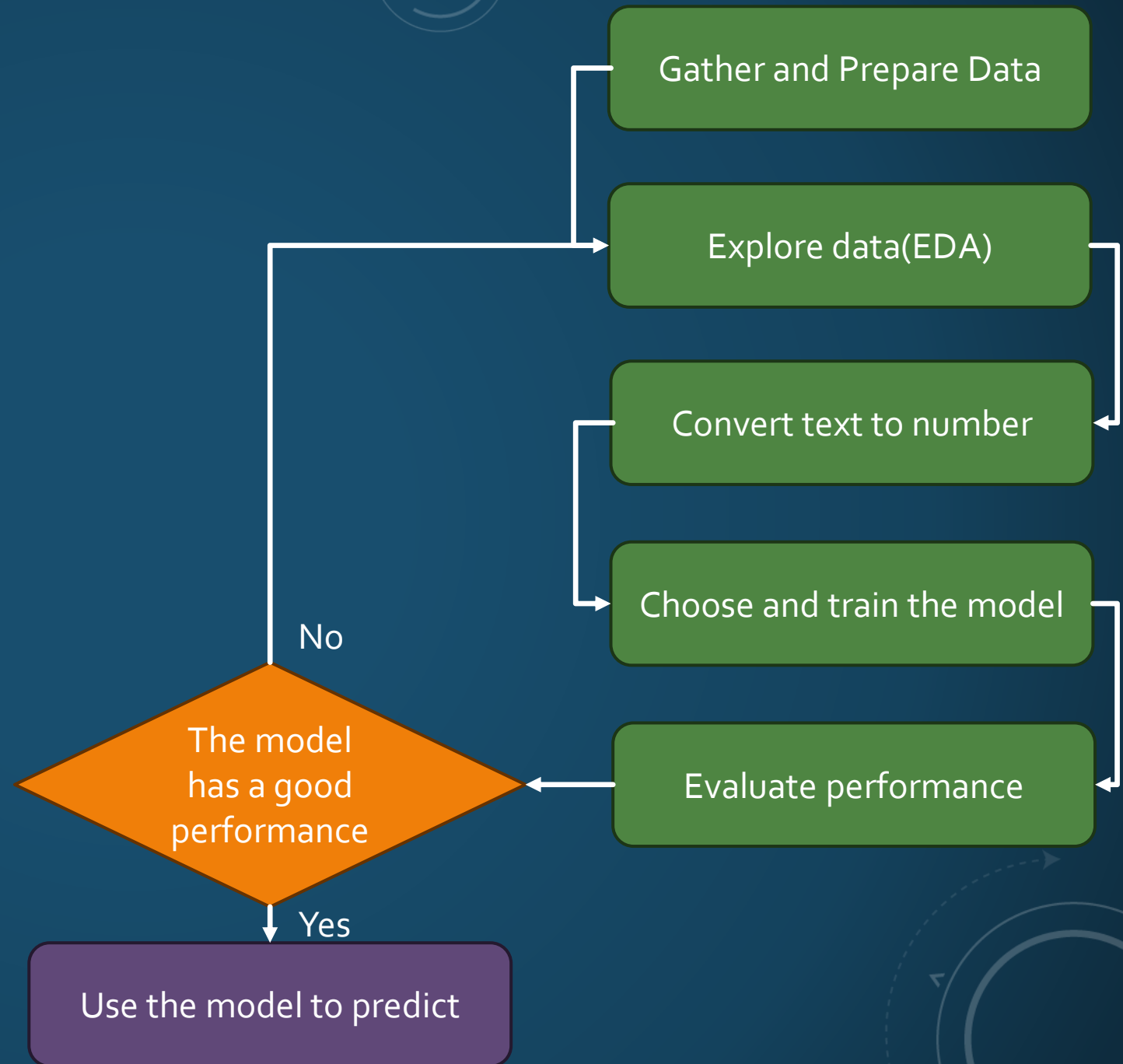
WHAT IS NLP?

Imagine you're talking to a friend. You understand their words, their jokes, and their feelings. **NLP (Natural Language Processing)** is the part of machine learning that tries to teach computers to do the same thing.

Its goal is to **make computers understand and use human language.**

RUN NLP STEP BY STEP

```
Gather and Prepare Data
        │
        ▼
Explore data(EDA)
        │
        ▼
Convert text to number
        │
        ▼
Choose and train the model
        │
        ▼
Evaluate performance
        │
        ▼
```

No ──► The model has a good performance
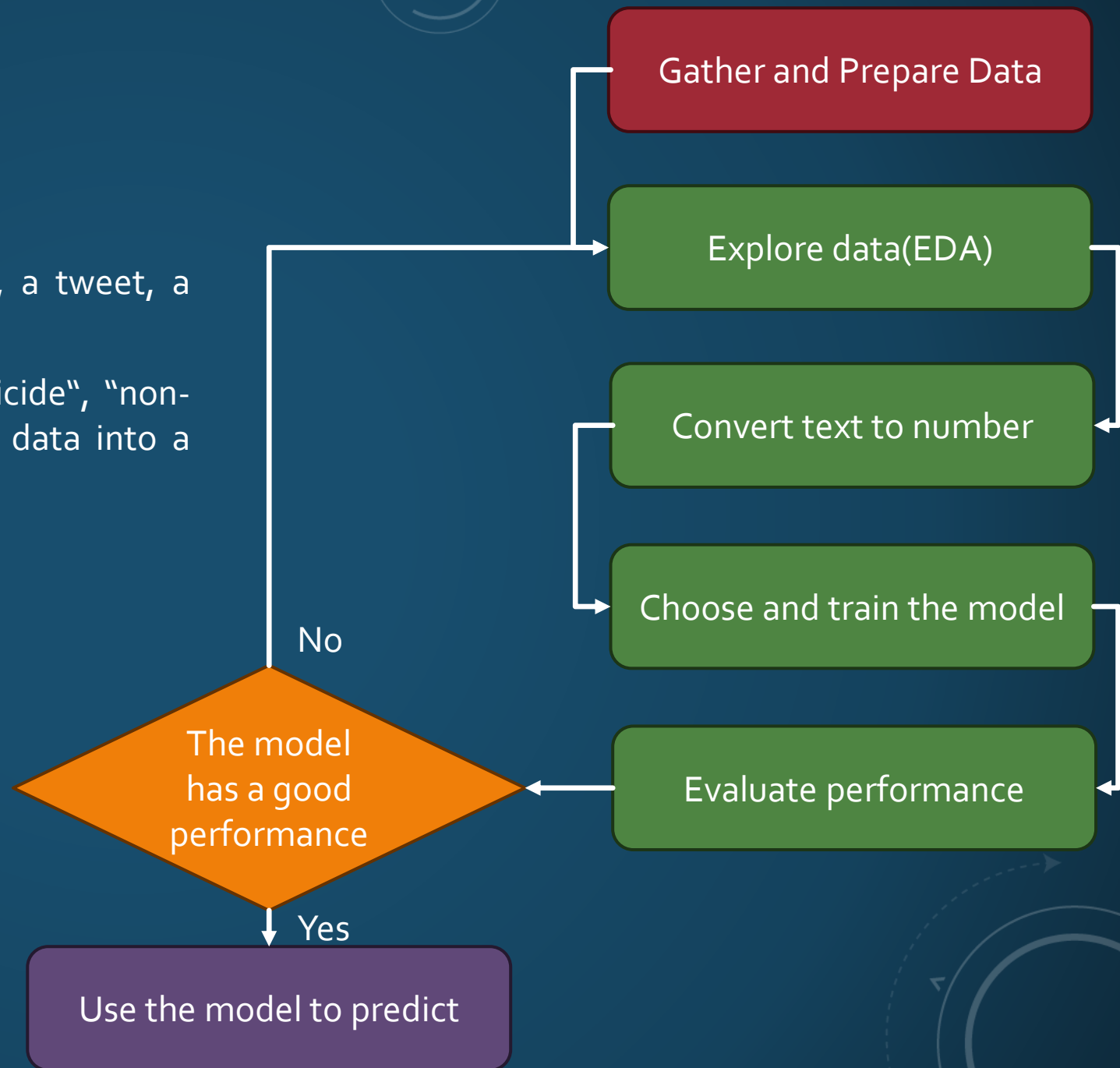
Yes ──► Use the model to predict

RUN NLP STEP BY STEP
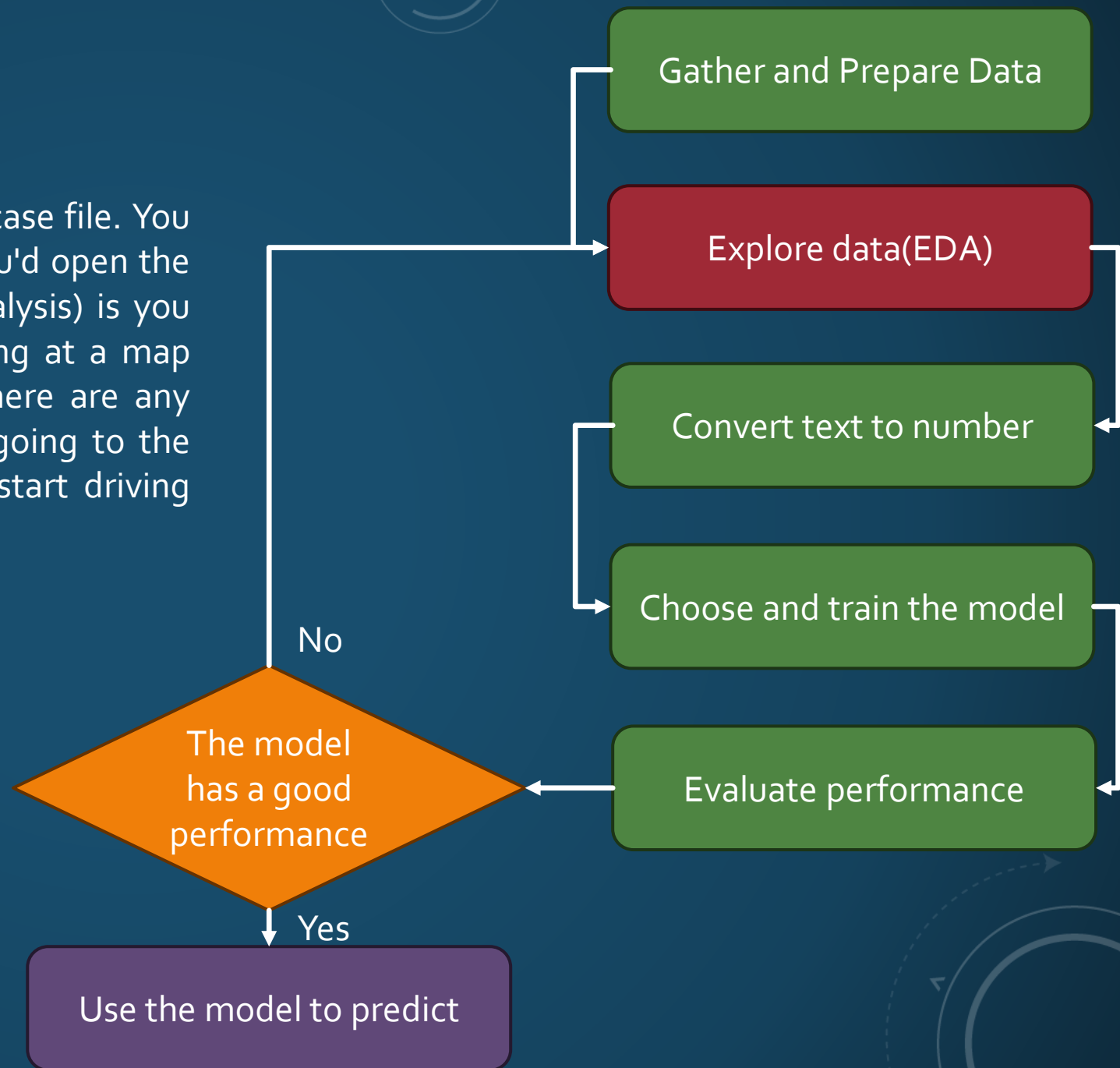
A dataset with at least two columns.

text: The raw text data (e.g., an email message, a tweet, a product description).

class: The category you want to predict (e.g., "suicide", "non-suicide"). CSV format is recommended. Load this data into a tool like Python (using libraries like Pandas).

Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

No

The model has a good performance
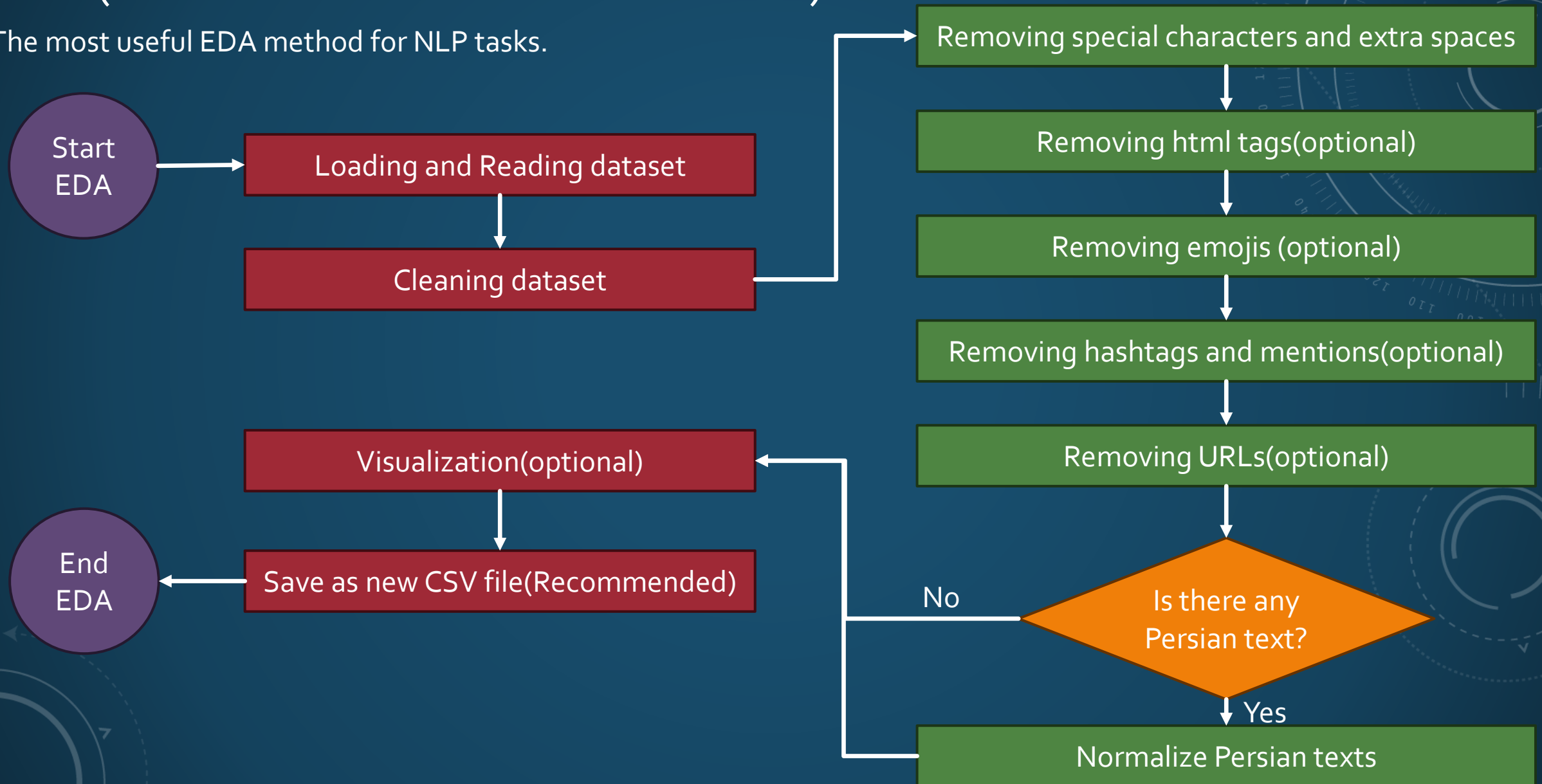
Yes

Use the model to predict

RUN NLP STEP BY STEP

Imagine you're a detective and you just got a new case file. You wouldn't just run to make an arrest, right? First, you'd open the file and look for clues. EDA (Exploratory Data Analysis) is you being a detective for your data. EDA is like looking at a map before a road trip. You check the route, see if there are any roadblocks (missing data), and make sure you're going to the right city (solving the right problem). If you just start driving without looking, you'll probably get lost!
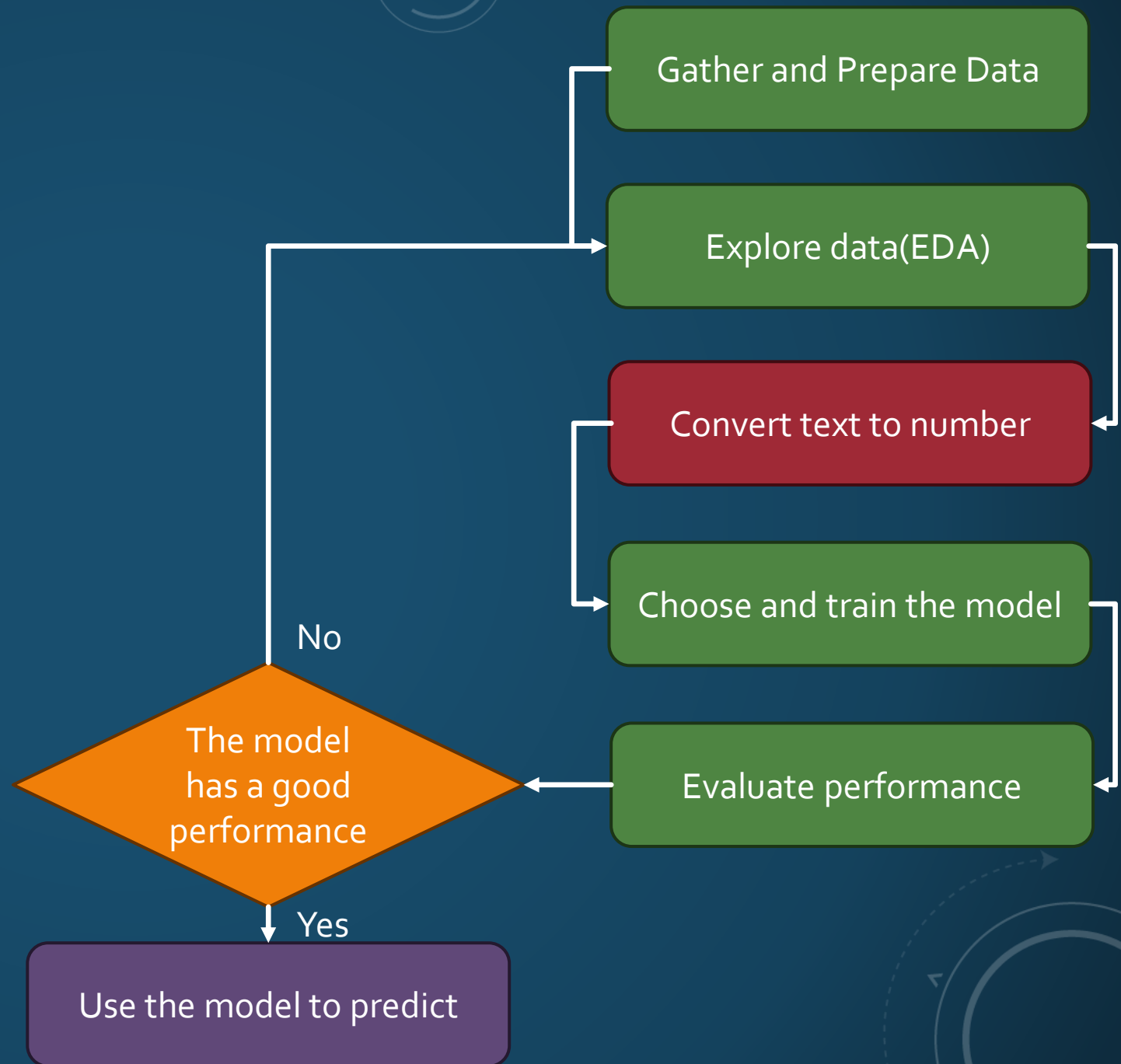
Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

The model has a good performance

No

Yes

Use the model to predict

EDA (EXPLORATORY DATA ANALYSIS)

The most useful EDA method for NLP tasks.

Start EDA → **Loading and Reading dataset** → **Cleaning dataset** → **Removing special characters and extra spaces** → **Removing html tags(optional)** → **Removing emojis (optional)** → **Removing hashtags and mentions(optional)** → **Removing URLs(optional)** → **Is there any Persian text?**

- No → **Visualization(optional)**
- Yes → **Normalize Persian texts** → **Visualization(optional)**

Visualization(optional) → **Save as new CSV file(Recommended)** → **End EDA**

RUN NLP STEP BY STEP

ML algorithms work with numbers

Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

No

The model has a good performance

Yes

Use the model to predict

RUN NLP STEP BY STEP

ML algorithms work with numbers

$$TF(t,d) = \frac{(Number\ of\ occurrences\ of\ term\ t\ in\ document\ d)}{(Total\ number\ of\ terms\ in\ the\ document\ d)}$$

$$IDF(t,D) = \log_e \frac{(Total\ number\ of\ documents\ in\ the\ corpus)}{(Number\ of\ documents\ with\ term\ t\ in\ them)}$$
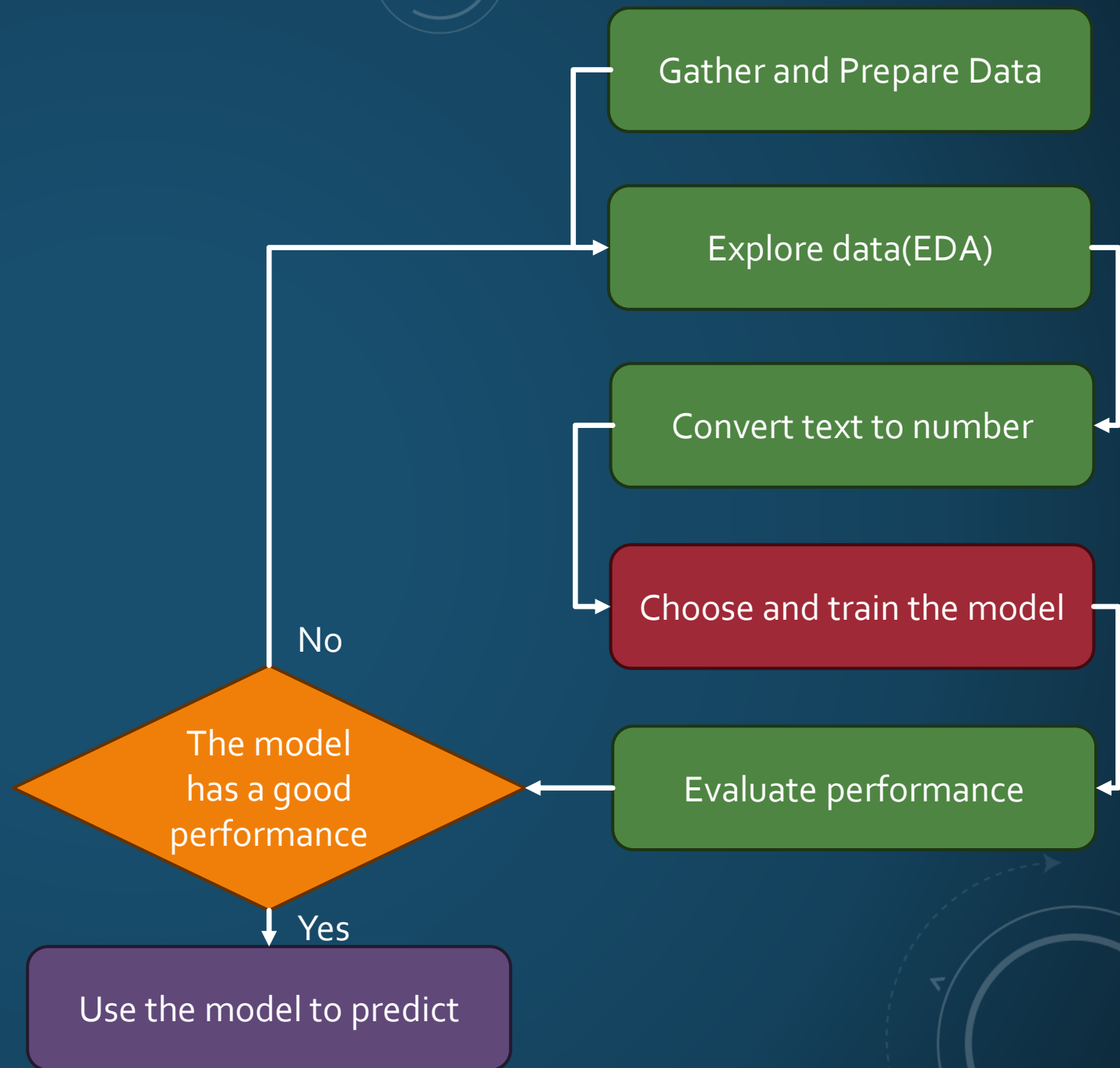
$$TF\text{-}IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

RUN NLP STEP BY STEP

Features: $X(x_1, x_2, ..., x_n)$

classes: C

Very hard

$$Pr[C | (x_1, x_2, ..., x_n)]$$

Gather and Prepare Data
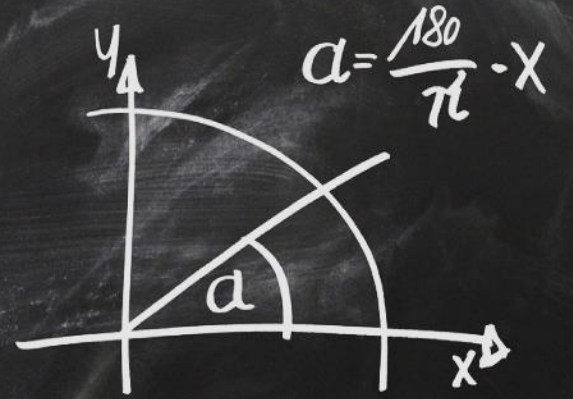
Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

The model has a good performance

No

Yes

Use the model to predict

NAÏVE BAYES CLASSIFIER

Thomas Bayes

NAÏVE BAYES CLASSIFIER

Types:

Gaussian Naive Bayes

Multinomial Naive Bayes

Bernoulli Naive Bayes

NAÏVE BAYES CLASSIFIER

Types:

Gaussian Naive Bayes

Multinomial Naive Bayes

Bernoulli Naive Bayes

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \, e^{-\frac{(v - \mu_k)^2}{2\sigma_k^2}}$$

NAÏVE BAYES CLASSIFIER

Types:

Gaussian Naive Bayes

Multinomial Naive Bayes

Bernoulli Naive Bayes

$$p(\mathbf{x} \mid C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{x_i}$$

NAÏVE BAYES CLASSIFIER

Types:

Gaussian Naive Bayes

Multinomial Naive Bayes

Bernoulli Naive Bayes

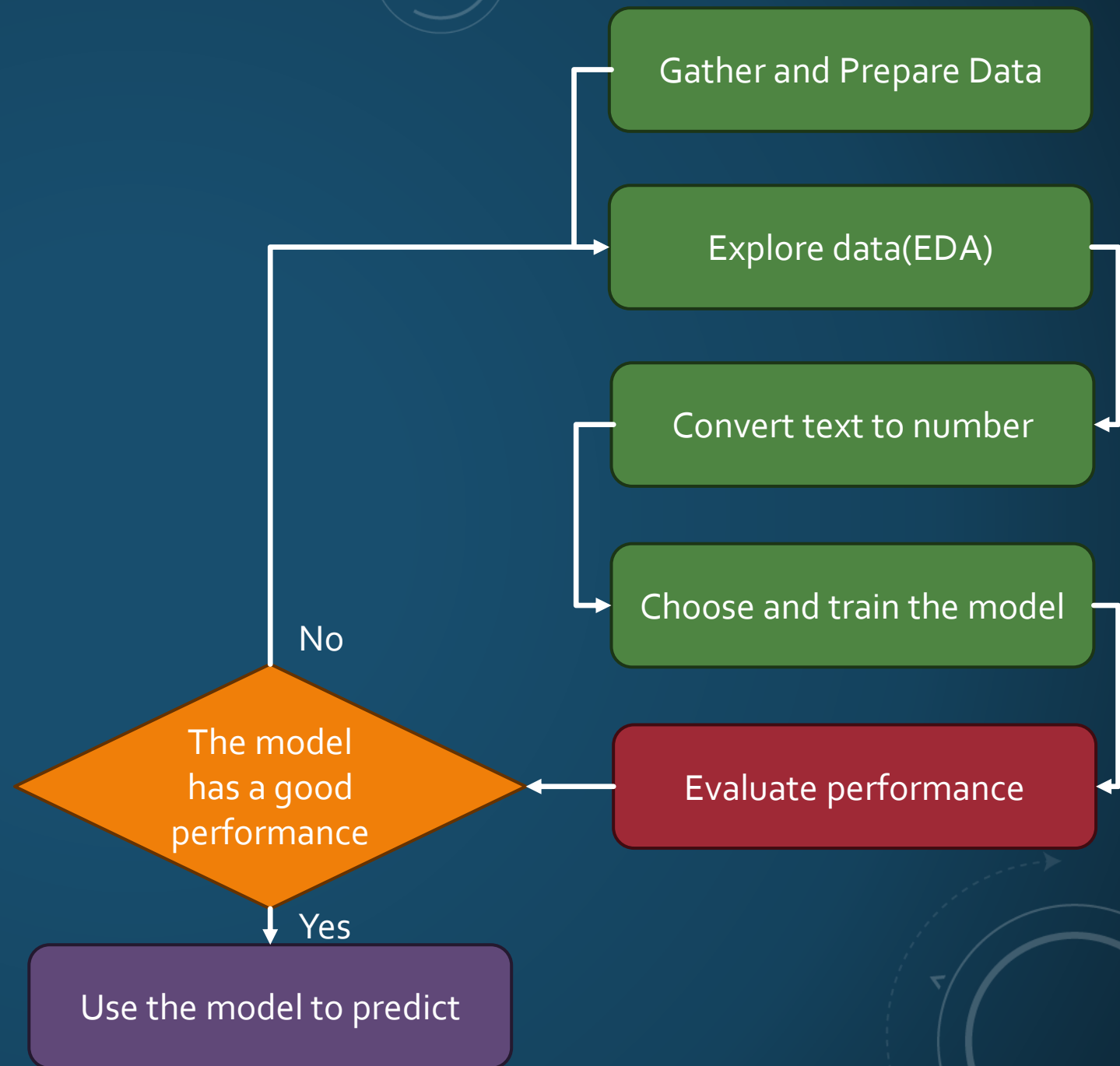$$p(\mathbf{x} \mid C_k) = \prod_{i=1}^{n} p_{ki}^{x_i}(1 - p_{ki})^{(1-x_i)}$$
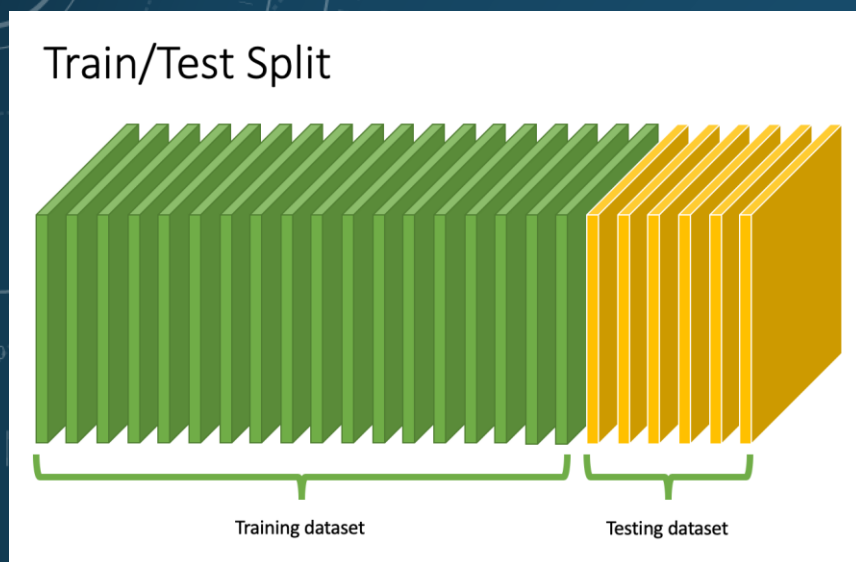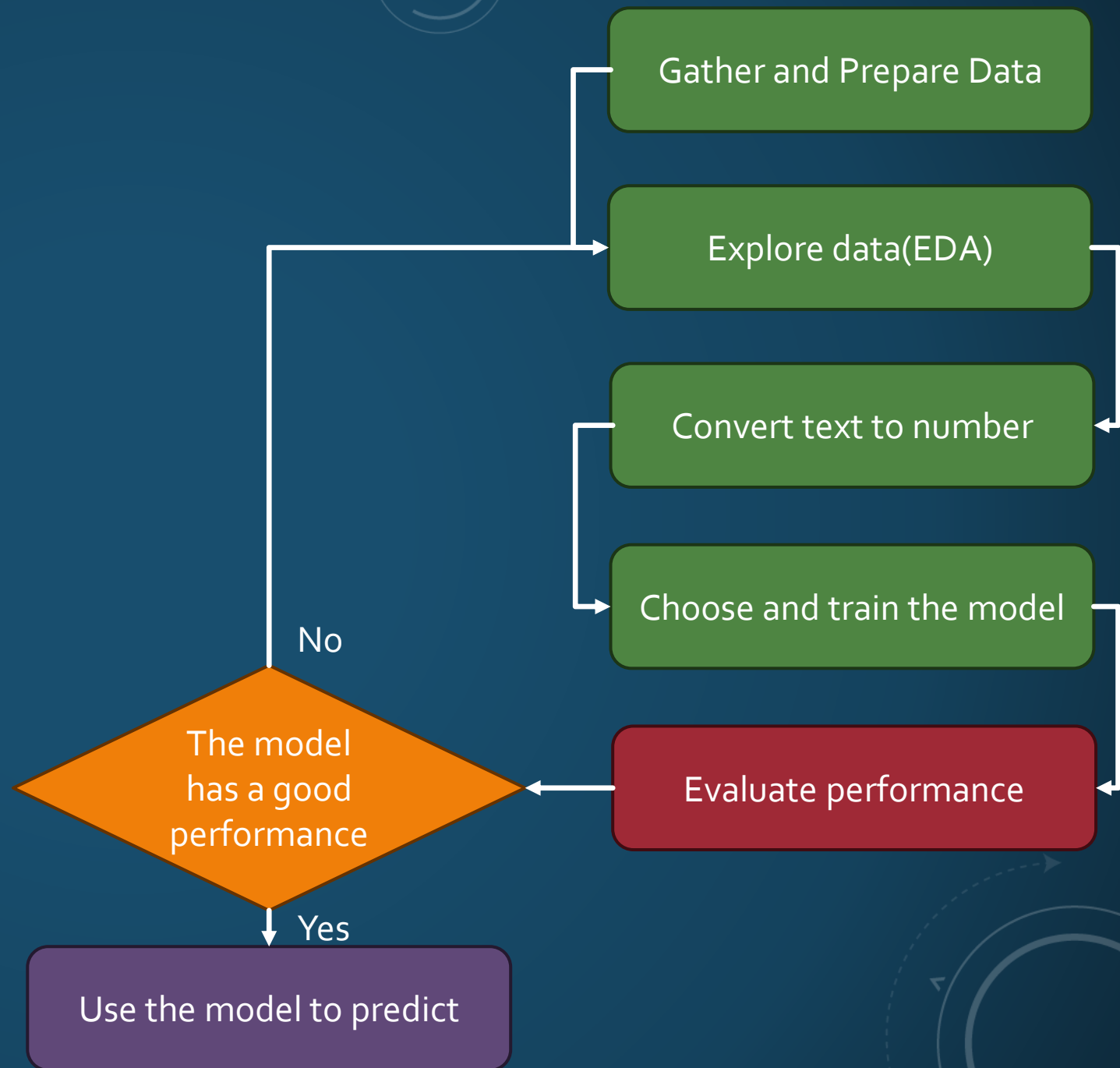
Which ne is better?

You can use the Voating classifier to find the best type according to your dataset model

RUN NLP STEP BY STEP

Train/Test Split

Training dataset Testing dataset

Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

No

The model has a good performance

Yes

Use the model to predict

RUN NLP STEP BY STEP

Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

Evaluate performance

The model has a good performance

No

Yes

Use the model to predict

RUN NLP STEP BY STEP

Positive | Negative

Predicted Label

| True Positive (TP) | False Positive (FP) | Positive |
| False Negative (FN) | True Negative (TN) | Negative |

True Label

Formulae for Classification Metrics

Precision $\dfrac{tp}{tp + fp}$

Recall (Sensitivity) $\dfrac{tp}{tp + fn}$

F1-score $\dfrac{2 * precision * recall}{precision + recall}$

Gather and Prepare Data

Explore data(EDA)

Convert text to number

Choose and train the model

No

The model has a good performance

Evaluate performance

Yes

Use the model to predict
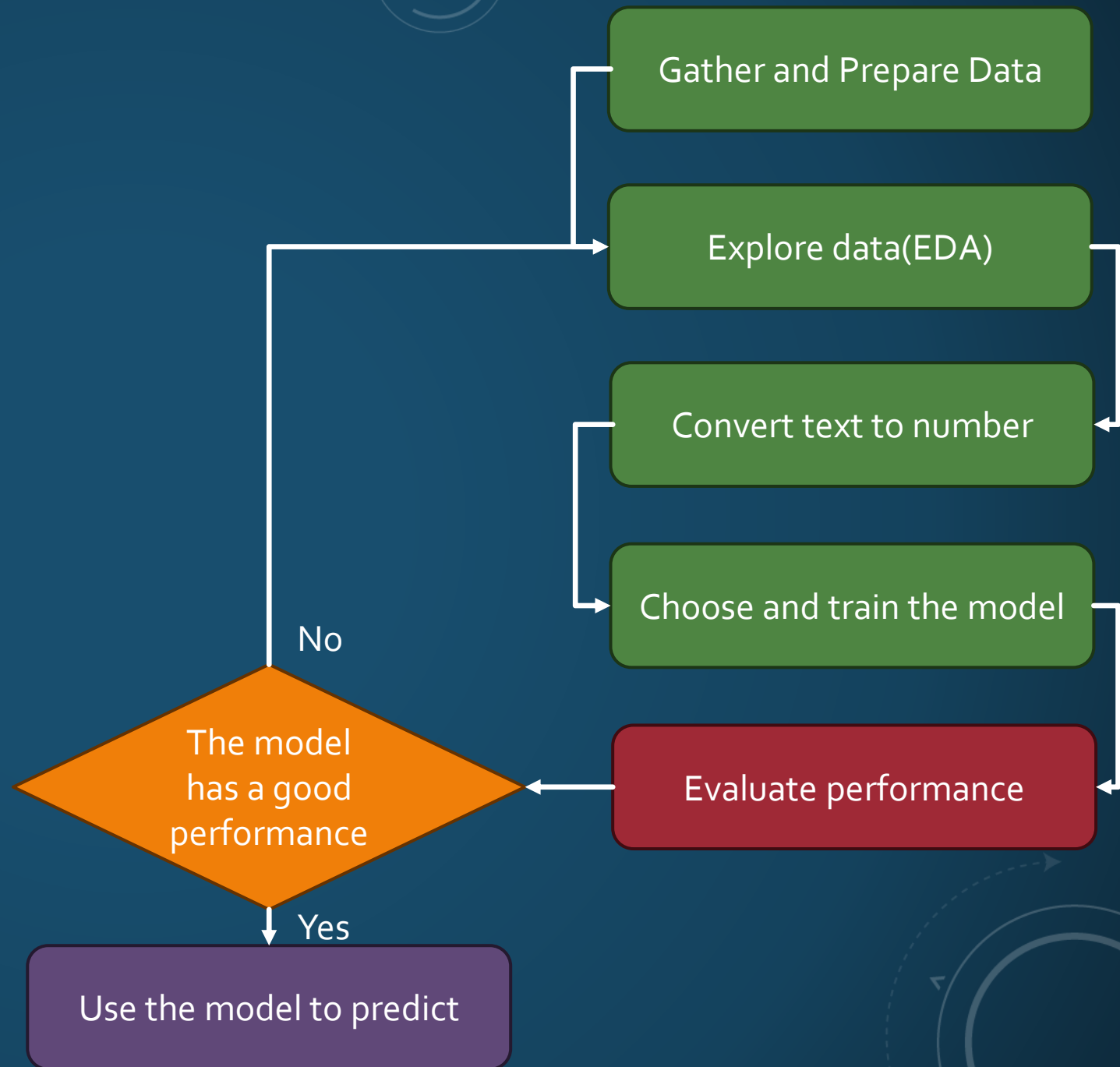
Use the model to predict

Input : I can't go on anymore. I'm tired and I want to disappear forever.
Output : 1
Input : I had a great day today and I got everything done on time.
Output : 0
Input : I'd better make dinner so I can get to my lessons later.
Output : 0
Input : I wish I never existed. Life is empty and meaningless.
Output : 1

IF YOU HAVE ANY QUESTIONS OR SUGGESTIONS, YOU CAN CONTACT ME.

Sajjadranjbaryazdi@gmail.com
https://www.linkedin.com/in/sajjad-ranjbar-yazdi-52aaa91a8
https://github.com/phantomf4321