



# Predict weather

**rain or snow**

By: Mada Abudahish & Samaher AL-hadi



# Background



this project is about historical weather around Szeged, Hungary , from 2006 to 2016, we build model that predict will it rain or snow.



## Dataset

Public source from Kaggle.



## Size

- 96000 records
- 12 columns



## Target

PRECIP TYPE



# Methodology



01

**Understanding the data**

02

**Gathering data**

03

**Exploratory Data Analysis**

04

**Data preparation**

05

**Classification models**

# Technologies and libraries



*NumPy*



jupyter



pandas



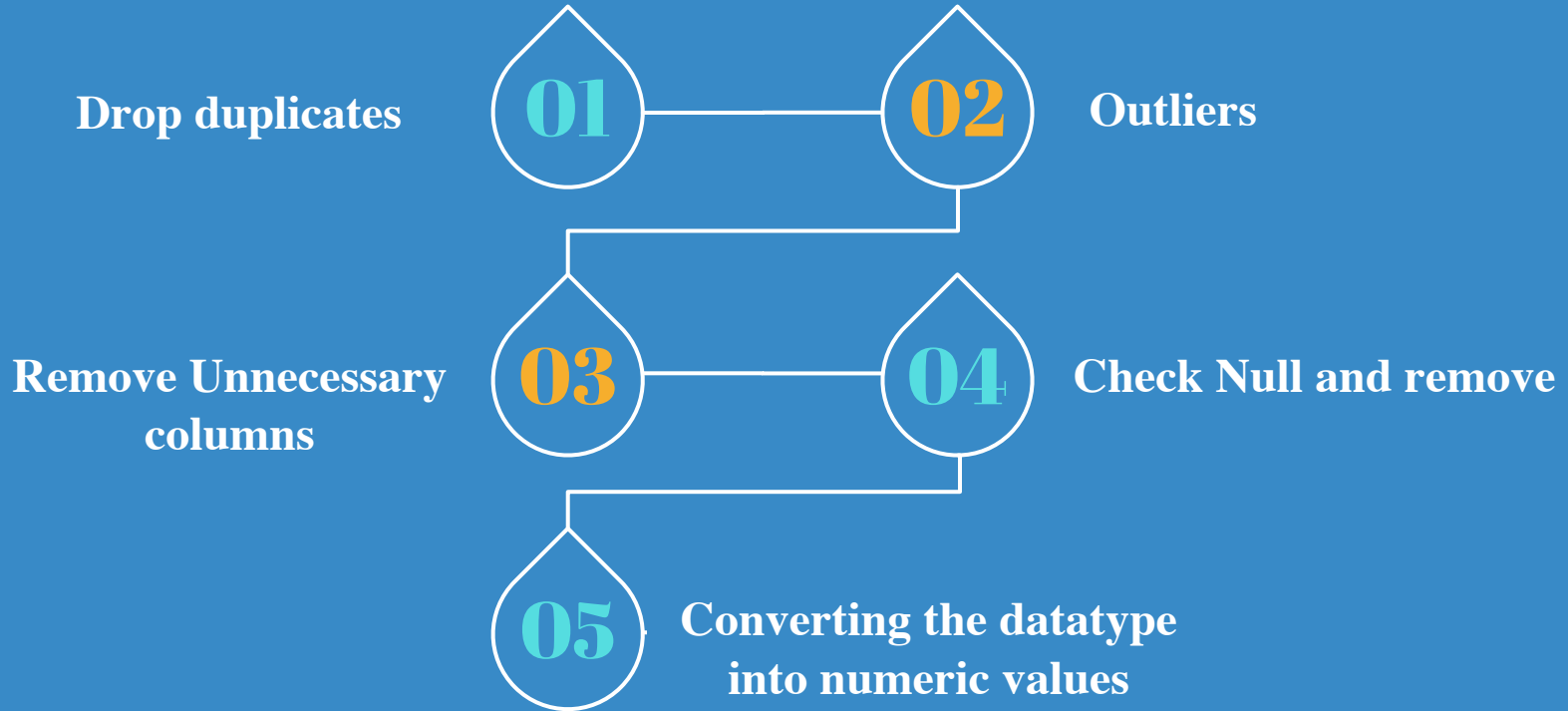
seaborn



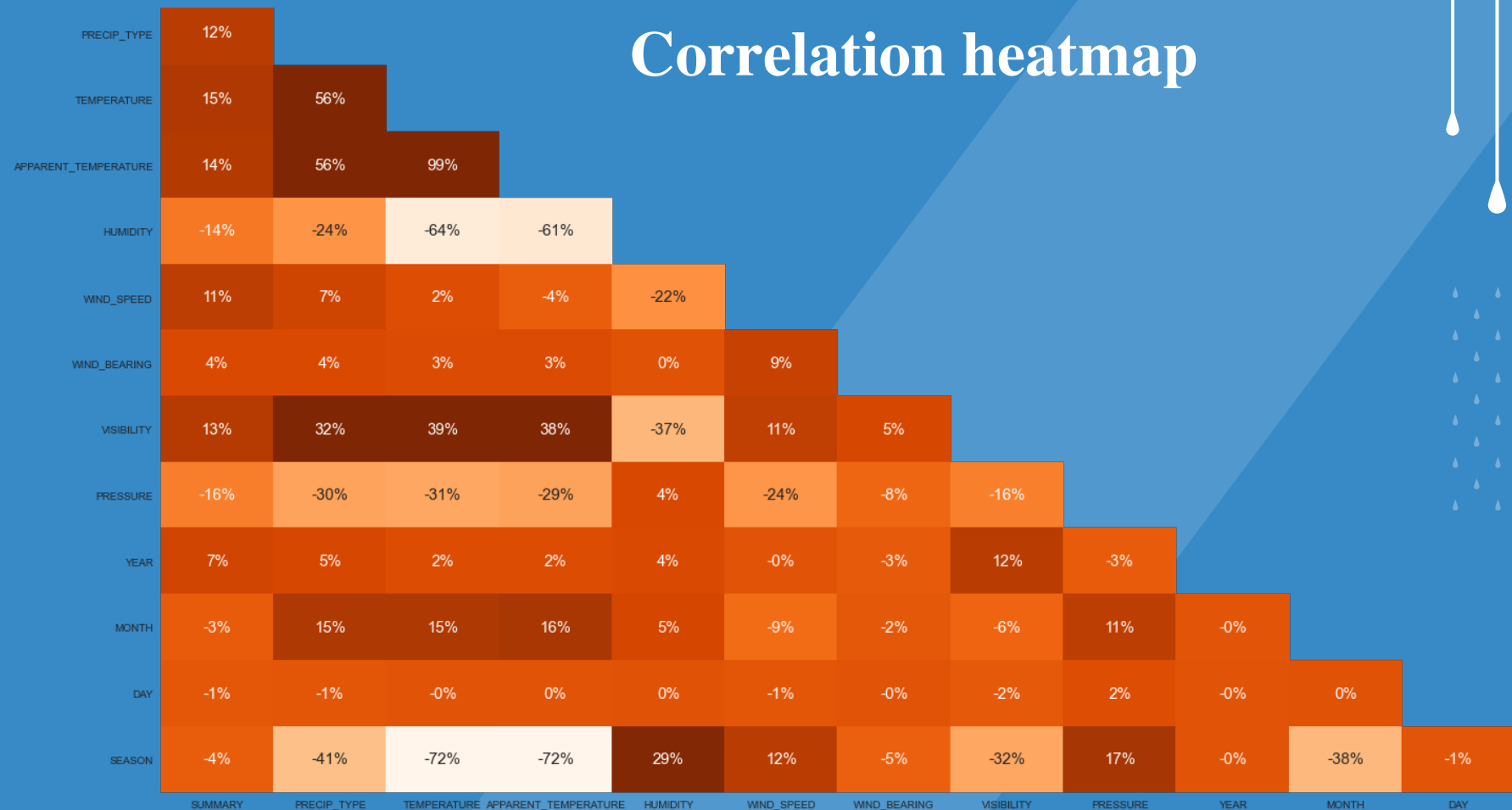
SQLAlchemy



# Exploratory Data Analysis



# Correlation heatmap



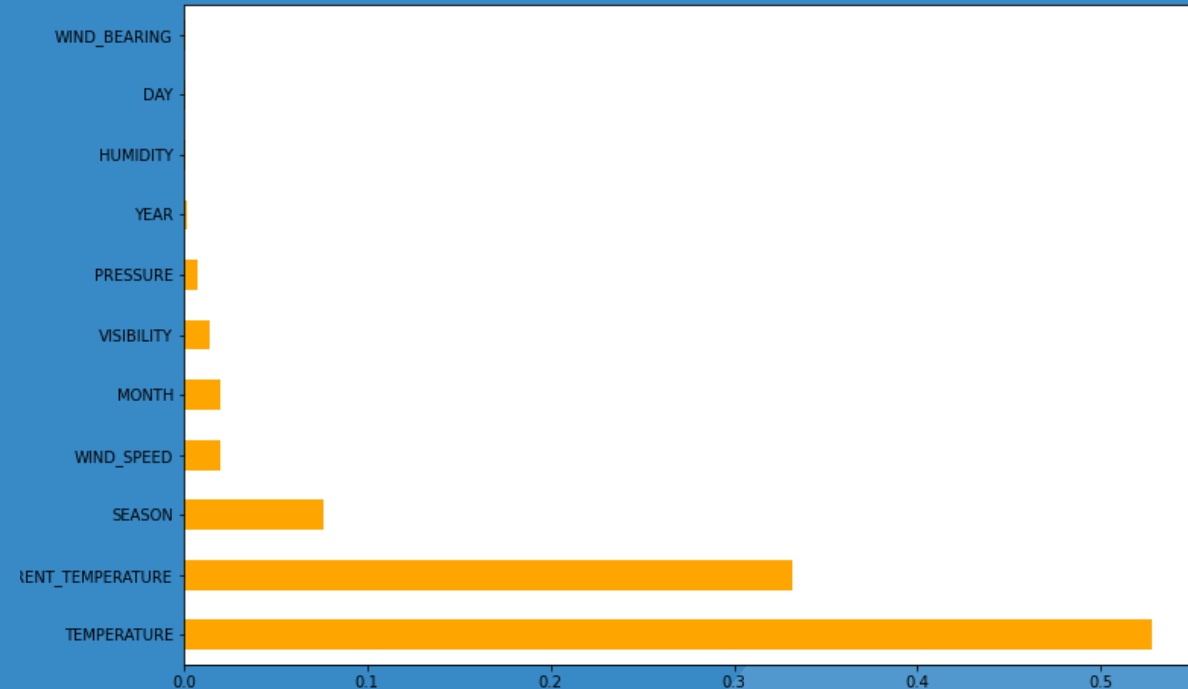
# feature importance



According to bar chart, these features:

1. **TEMPERATURE**
2. **APPARENT TEMPERATURE**
3. **SEASON**
4. **WIND SPEED**
5. **VISIBILITY**
6. **MONTH**
7. **PRESSURE**
8. **YEAR**

Are the most important features in weather predict.



# Data Preparation

## Feature Selection

- Drops both the 'loud cover' and the 'week\_day' columns

## Feature Engineering

- Add new column ('SEASON').
- Encoding the columns into categorical values.
- Scaling using standard Scaler.

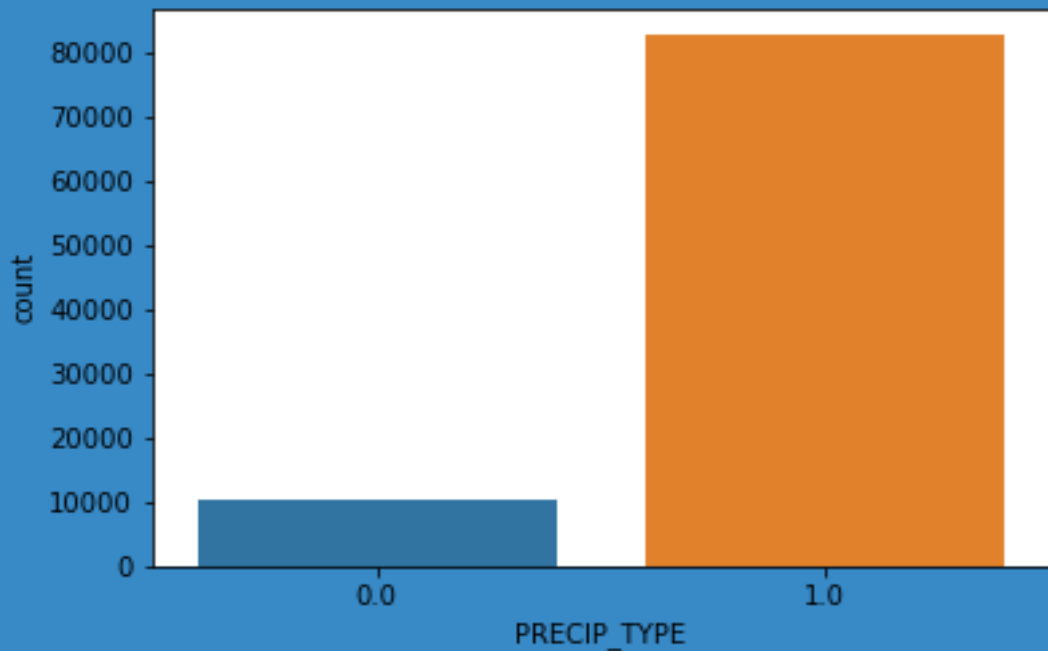
## Imbalanced dataset

- SMOTE was use for handling the imbalanced

## Voting Classifier



# Imbalance



There were snow (11.102%) and rain (88.898%).



**SMOTE**

**SNOW: 83010**

**RAIN: 83010**



**ADA**

**SNOW: 81040**

**RAIN: 81348**

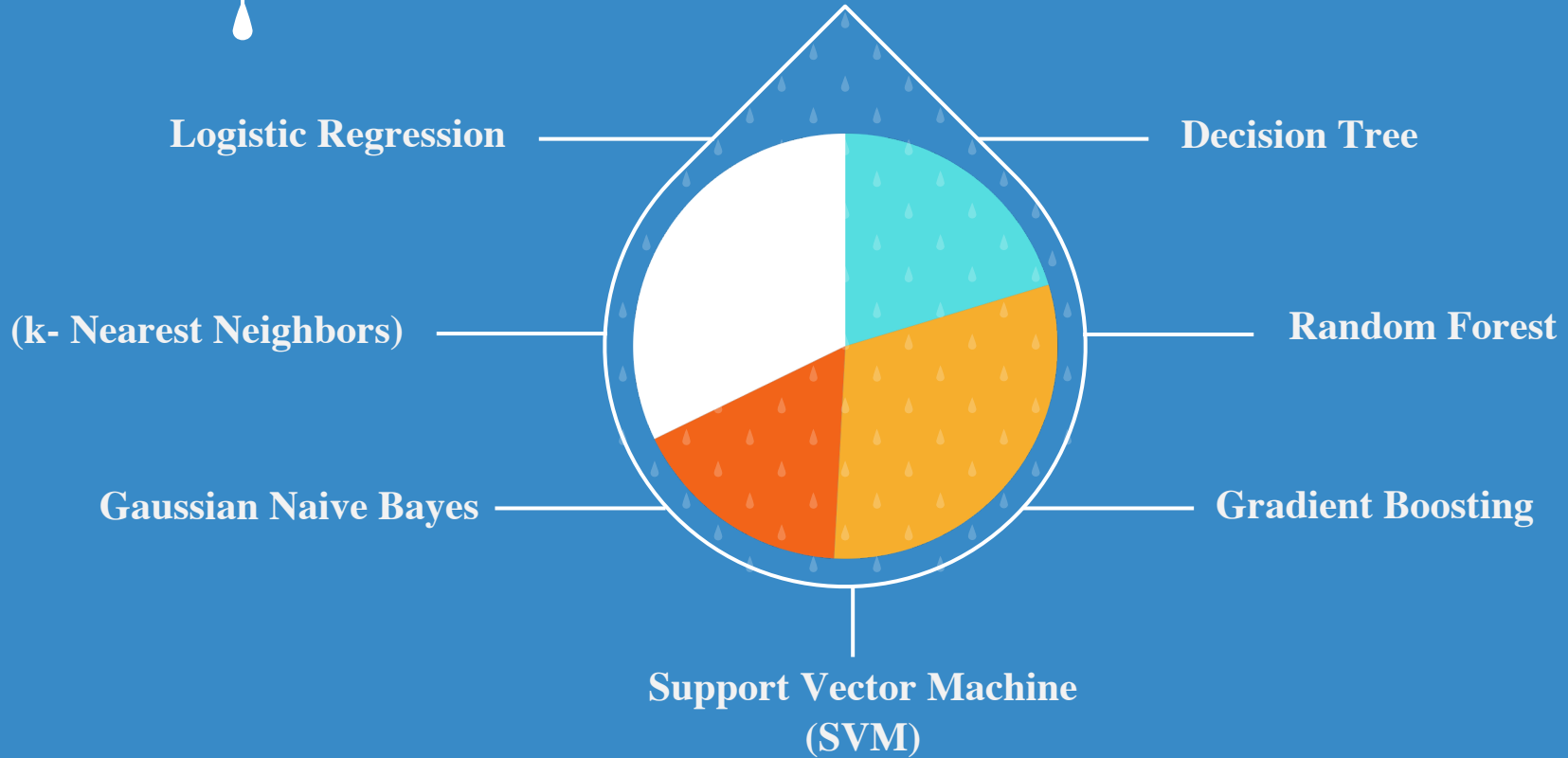


**RANDOM OVER  
SAMPLER**

**SNOW: 81348**

**RAIN: 81348**

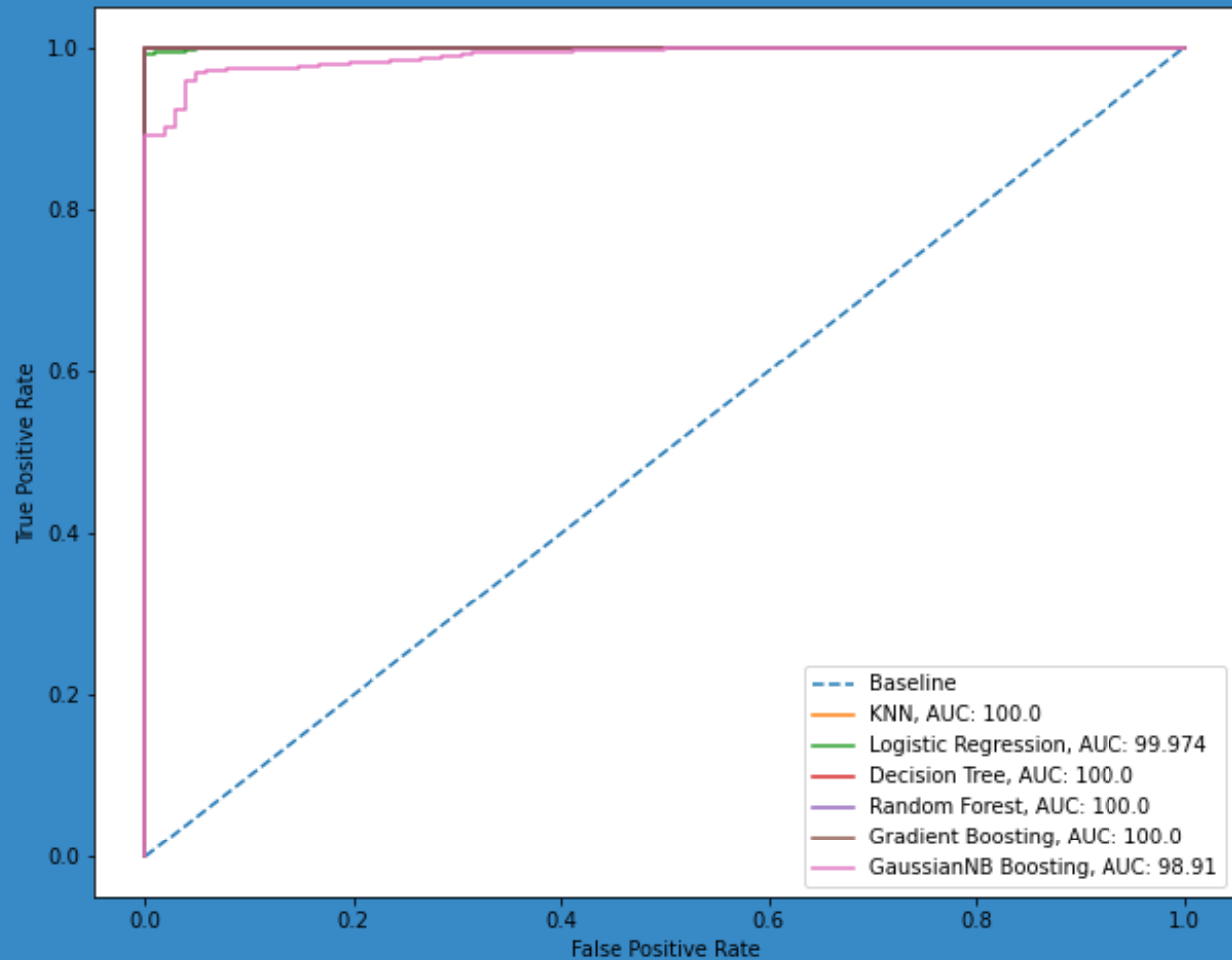
# Classification models





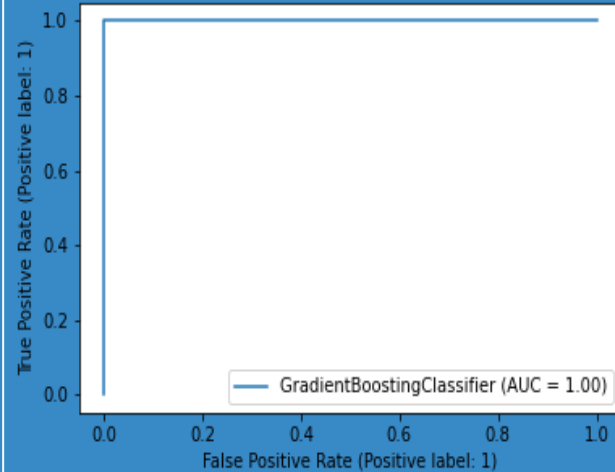
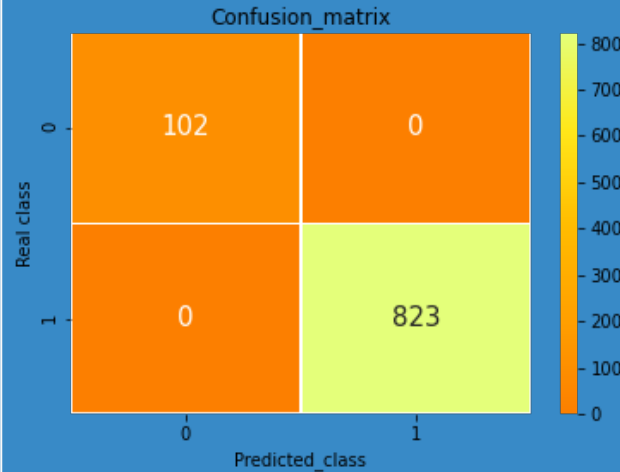
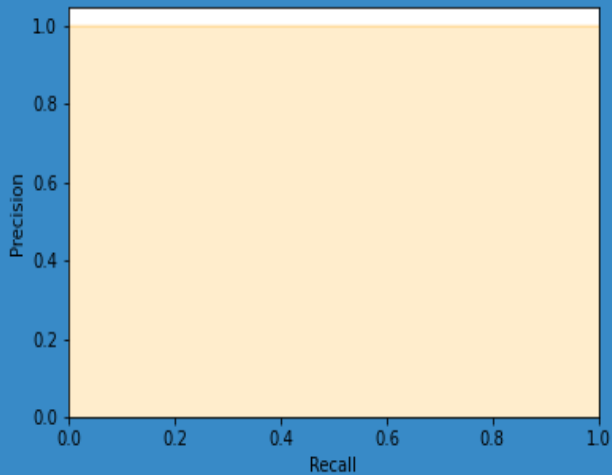
# Classification models

	Precision	Recall	F1	Accuracy
Logistic Regression	1.00	0.9927	0.9963	0.9935
Nearest Neighbors	0.9939	0.9891	0.9915	0.9848
Decision Tree	1.00	1.00	1.00	1.00
Gradient Boosting	1.00	1.00	1.00	1.00
Random Forest	1.00	1.00	1.00	1.00

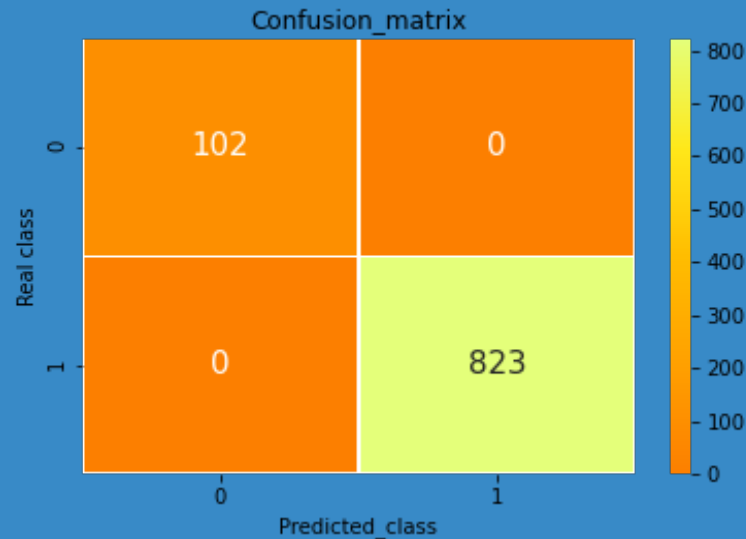
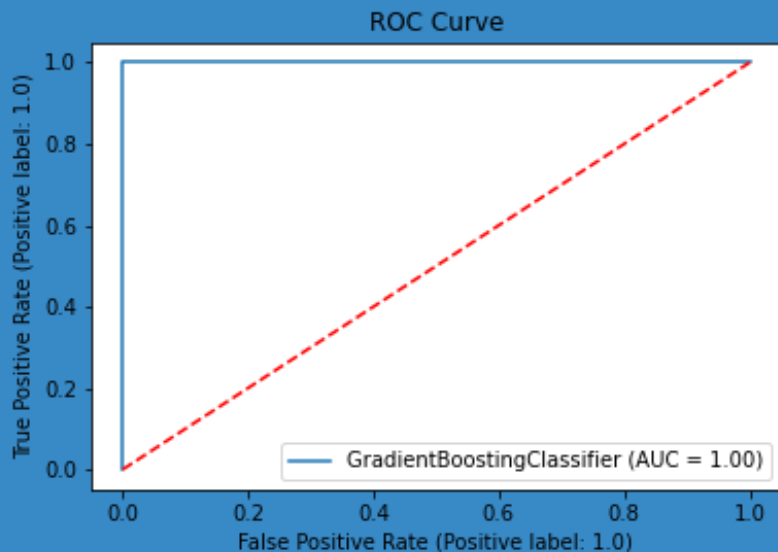


# ROC Curve

# Gradient Boosting Model



# Gradient Boosting (test)



**Precision**

**Recall**

**F1**

**Accuracy**

**Gradient  
Boosting**

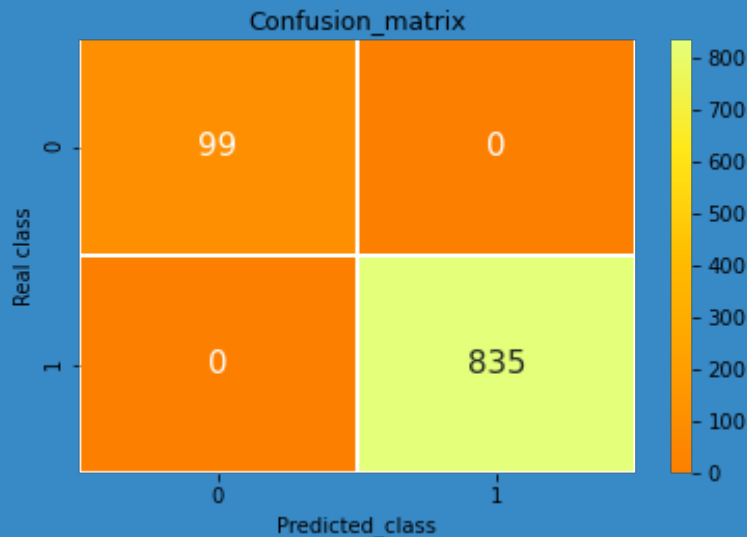
**1.0**

**1.0**

**1.0**

**1.0**

# Voting Classifier (test)



	Precision	Recall	F1	Accuracy
Voting Classifier	1.0	1.0	1.0	1.0





# Future Works

- **Another dataset / Scraped dataset**
  - **Feature Engineering (add new column)**
- 
- 



THANKE YOU !