

Classification of Stingless Bee Honey

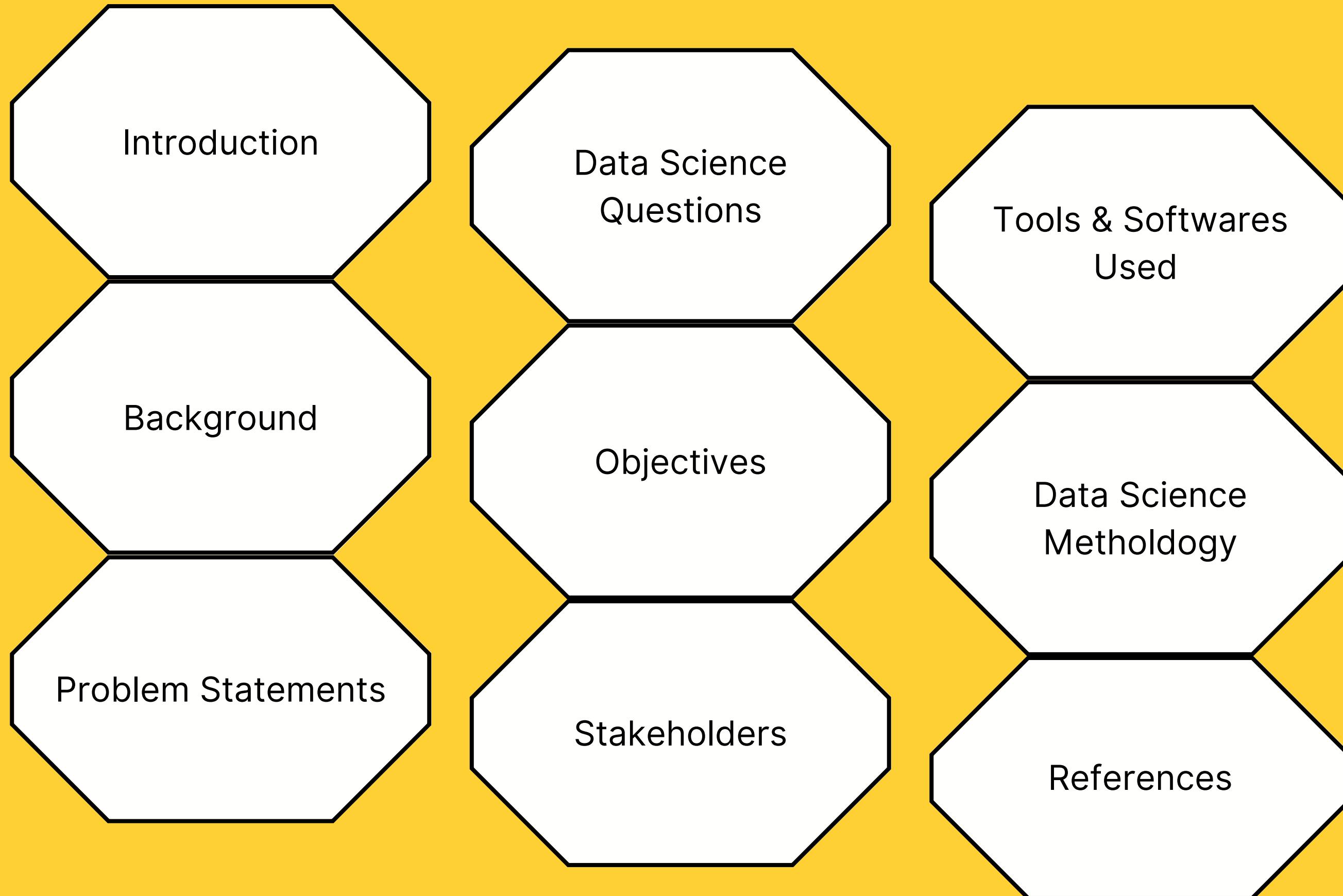


Data Science Project

Prepared by Muhammad Zariff Wafiy

Supervisor : Dr.Nor Liyana Shuib

OVERVIEW



INTRODUCTION



Stingless Bee, *Meliponini*

Stingless bee genus contains 500 species, most of which are found in Latin America, mainland Australia, Africa, Eastern and Southern Asia, and the Middle East (Abd Jalil, 2017)

Very crucial to the environment, the economy, and culture with the role of the primary pollinators for a variety of tropical wild and domesticated plants





As Antioxidant

Phenolic acids found in stingless bee honey, potent antioxidant that can accelerate the healing of wounds by promoting the proliferation of new cells.



As Antimicrobial

Non-peroxide activity affects the antibacterial impact of stingless bee honey that is responsible for preventing microbial growth.



As Anti-inflammatory

An enzyme named 5-LOX which is necessary for the production of proinflammatory mediators, is inhibited by the cerumen of honeys.

BACKGROUND

While there are numerous researches conducted internationally, further researches specifically in the domain of local stingless bee honey are needed to strengthen the local commercial industry of biological consumables in terms of purity and authenticity. Additionally, there are still no standards of general practise and/or quality for this kind of honey. (Pimentel et al., 2021).

This project aims to deliver an interactive dashboard which could assist potential customers of stingless bee honeys.



PROBLEM STATEMENTS



Influence of chemical composition in stingless bee honey

Unsure of the correlation between chemical compositions of honey and the antioxidant activities in samples.



Acknowledge regarding different species of stingless bee honey

The physicochemical and antioxidant characteristics of stingless bee honey were different from those of *Apis mellifera* honey. (Shamsudin et al., 2019)

DATA SCIENCE QUESTIONS



1

How different antioxidant activities of honey can affect chemical composition in them?

2

What does each species of stingless bee contain that differs from one another?

Objective 01

To identify the correlation between different chemical compositions with different antioxidant activities.

Objective 02

To develop a classification model to classify species of stingless bee based on samples of honey.

OBJECTIVES

Objective 03

To deploy the classification as an interactive application that allows user to predict species of stingless bees based on honey attributes.

STAKEHOLDERS



**Stingless bee
honey consumers**



**Scientists &
Researchers**

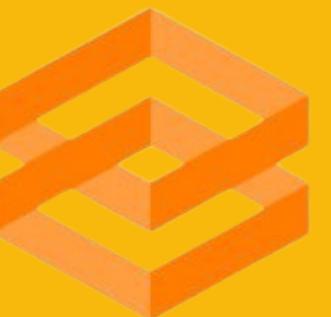
TOOLS & SOFTWARE USED



Programming
Language



Programming
Platform



gradio

Graphical User
Interface

Data Collection

Dataset made available by stingless bee honey research team.

Exploratory Data Analysis

Summarizes of the whole dataset through statistical graphics

Data Preprocessing

Includes data cleaning, and data reduction.

Descriptive Analytics

Plot several graphs to analyze correlations between variables

DATA SCIENCE METHODOLOGY

Model Deployment

Develop an interactive graphical user interface to deploy classification model.

Model Evaluation

Have the best model with the highest accuracy set as the main model for deployment.

Predictive Modelling

Build different classification models to predict species of stingless bee honey.

DATA COLLECTION



Data collected from the Centre for Natural Products Research and Drug Discovery (CENAR) research team.

Contains structured data of 158 samples of stingless bee honeys from across 5 state in Malaysia

A	B	C	D	E	F	G	H	I	J	K
1	No	Sample	PH	TPC ($\mu\text{g GA}$)	TFC ($\mu\text{g QE}$)	ABTS (%) in FRAP ($\mu\text{g F}$)	Water con	Fructose c	Glucose cc	Remark
2	1	A itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9
3	2	A itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2
4	3	A itama iii	3	152.83	44.57	15.21	14.63	30.5	69.7	69.4
5	4	A itama iv	3.2	203.06	43.97	20.4	20.21	28.8	71.5	71.3
6	5	A itama v	3.3	178.09	44.59	17.5	16.19	26.1	74.3	74.1
7	6	A thoracic	3	156.04	35.08	11.57	8.57	30	70.6	70.4
8	7	A thoracic	2.9	149.63	38.56	12.25	9.63	30.1	70.1	69.8
9	8	A thoracic	2.9	139.86	37.25	11.28	7.02	30.9	69.3	69
10	9	A thoracic	3.1	164.17	39.59	13.38	11.23	30	70.2	70
11	10	A thoracic	3.1	148.37	37.52	12.61	11.43	30.2	70	69.8
12	11	B itama i	3	164.75	58.27	12.95	21.23	28.3	72.4	72.1
13	12	B itama ii	3	156.89	53.73	13.86	13.82	26.8	73.6	73.4
14	13	B itama iii	2.9	149.51	45.23	11.42	8.97	28	72.3	72.1
15	14	B itama iv	3.3	159.21	45.95	14.21	13.78	28.9	71.3	71.1
16	15	B itama v	3.3	189.66	52.88	16.34	18.92	26.1	74.3	74.1
17	16	B thoracic	3.3	139.3	32.26	11.42	8.2	26.3	72.4	73.7
18	17	B thoracic	3.2	141.62	32.48	12.57	6.43	25.6	73.6	74.5
19	18	B thoracic	3.2	134.36	31.9	10.66	7.04	27.6	72.3	72.4
20	19	B thoracic	3.4	148.45	32.67	12.94	11.83	27.1	73.2	73.1

SCIENTIFIC TERMINOLOGIES

Potential of Hydrogen, pH

Indicates if the sample is acidic or basic.

Total Phenolic Content, TPC

Chemical compound.

Total Flavonoid Content, TFC

Chemical compound.

2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid, ABTS

To test antioxidant activity.

Ferric Reducing Antioxidant Power , FRAP

To test antioxidant activity.

Water Content

Moisture level present in sample.

Fructose Content

Content of sugar (Fructose) in sample.

Glucose Content

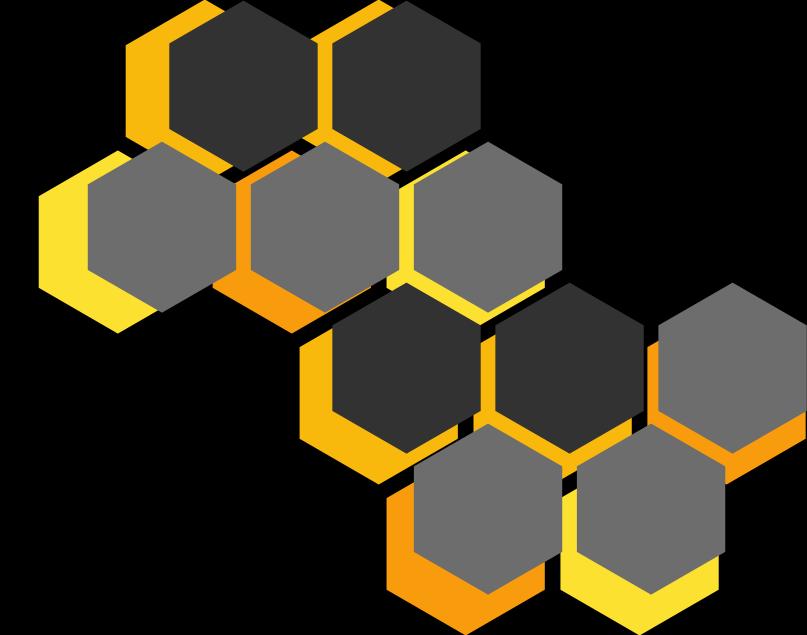
Content of sugar (Glucose) in sample.

EXPLORATORY DATA ANALYSIS

No	Sample	PH	TPC (µg GAE/mg honey)	TFC (µg QE/mg honey)	ABTS (% inhibition)	FRAP (µg FeSO4/mg honey)	Water content (% weight/weight)	Fructose content (% weight/weight)	Glucose content (% weight/weight)	Remark
0	1	A itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9
1	2	A itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2
2	3	A itama iii	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4
3	4	A itama iv	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3
4	5	A itama v	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1

First five observations

EXPLORATORY DATA ANALYSIS



```
RangeIndex: 158 entries, 0 to 157
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype  
 ---  -- 
 0   No              158 non-null    int64  
 1   Sample          158 non-null    object  
 2   PH              158 non-null    float64 
 3   TPC (µg GAE/mg honey) 158 non-null    float64 
 4   TFC (µg QE/mg honey) 158 non-null    float64 
 5   ABTS (% inhibition) 158 non-null    float64 
 6   FRAP (µg FeSO4/mg honey) 158 non-null    float64 
 7   Water content (% weight/weight) 158 non-null    object  
 8   Fructose content (% weight/weight) 158 non-null    float64 
 9   Glucose content (% weight/weight) 158 non-null    float64 
 10  Remark          41 non-null     object  
dtypes: float64(7), int64(1), object(3)
```

Shows the overall overview of the dataset including null values and attribute data types.

Summary of the dataset

EXPLORATORY DATA ANALYSIS

```
array(['A itama i', 'A itama ii', 'A itama iii', 'A itama iv',
       'A itama v', 'A thoracica i', 'A thoracica ii', 'A thoracica iii',
       'A thoracica iv', 'A thoracica v', 'B itama i', 'B itama ii',
       'B itama iii', 'B itama iv', 'B itama v', 'B thoracica i',
       'B thoracica ii', 'B thoracica iii', 'B thoracica iv',
       'B thoracica v', 'C itama i', 'C itama ii', 'C itama iii',
       'C itama iv', 'C itama v', 'C thoracica i', 'C thoracica ii',
       'C thoracica iii', 'C thoracica iv', 'C thoracica v', 'D itama i',
       'D itama ii', 'D itama iii', 'D apicalis i', 'D apicalis ii',
       'D apicalis iii', 'D thoracica i', 'D thoracica ii',
       'D thoracica iii', 'D erythrogastera i', 'D erythrogastera ii',
       'D erythrogastera iii', 'D binghami i', 'D binghami ii',
       'D binghami iii', 'E itama i', 'E itama ii', 'E itama iii',
       'E itama iv', 'E itama v', 'E Binghami i', 'E Binghami ii',
       'E Binghami iii', 'E Binghami iv', 'E Binghami v', 'E Fimbriata i',
       'E Fimbriata ii', 'E Fimbriata iii', 'E Fimbriata iv',
       'E Fimbriata v', 'E Apicalis i', 'E Apicalis ii', 'E Apicalis iii',
       'E Apicalis iv', 'E Apicalis v', 'E Erythrogastera i',
       'E Erythrogastera ii', 'E Erythrogastera iii', 'E Erythrogastera iv',
```

A lot of unique values for sample due to iterations and state codes

Unique values of samples

DATA PREPROCESSING

DATA CLEANING STEPS

- 1 Renaming columns for better understanding
- 2 Separate sample extract location from sample name to a new column
- 3 Change the value of state codes to real state names
- 4 Remove unnecessary species iteration
- 5 Impute missing values
- 6 Change incorrect data types
- 7 Remove unnecessary / unused columns
- 8 Randomize dataset



RENAMEING COLUMNS

No	Sample	PH	TPC (µg GAE/mg honey)	TFC (µg QE/mg honey)	ABTS (% inhibition)	FRAP (µg FeSO4/mg honey)	Water content (% weight/weight)	Fructose content (% weight/weight)	Glucose content (% weight/weight)	Remark	
0	1	A itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9	NaN
1	2	A itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2	NaN
2	3	A itama iii	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4	NaN
3	4	A itama iv	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3	NaN
4	5	A itama v	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1	NaN

Before renaming columns



Sample_ID	Sample_Name	PH	TPC	TFC	ABTS	FRAP	Water	Fructose	Glucose	Remark	
0	1	A itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9	NaN
1	2	A itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2	NaN
2	3	A itama iii	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4	NaN
3	4	A itama iv	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3	NaN
4	5	A itama v	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1	NaN

After renaming columns

SPLIT INTO NEW COLUMNS

```

honey[["State", "Sample_Name"]] = honey.Sample_Name.str.split(" ", 1, expand=True)

```

Sample_ID	Sample_Name	PH	TPC	TFC	ABTS	FRAP	Water	Fructose	Glucose	Remark	State
0	1	itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9	NaN A
1	2	itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2	NaN A
2	3	itama iii	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4	NaN A
3	4	itama iv	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3	NaN A
4	5	itama v	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1	NaN A

After splitting sample name and state code

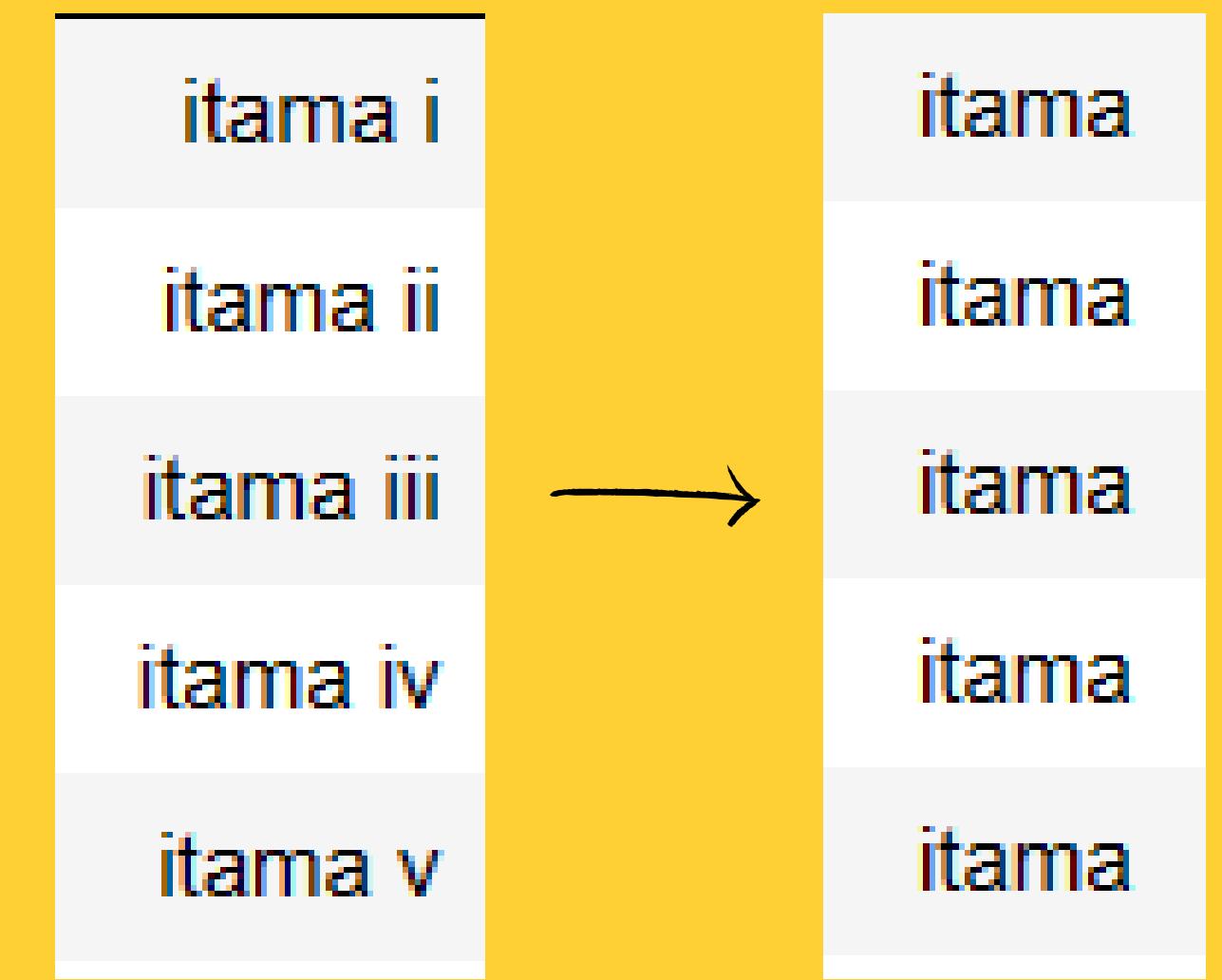
REPLACE VALUES

```
...  
honey["State"] = honey["State"].replace(["A", "B", "C"], "Kedah")  
honey["State"] = honey["State"].replace(["D", "E"], "Selangor")  
honey["State"] = honey["State"].replace(["F", "G", "H"], "Johor")  
honey["State"] = honey["State"].replace(["I", "J"], "Pahang")  
honey["State"] = honey["State"].replace(["K", "L"], "Terengganu")
```

Sample_ID	Sample_Name	PH	TPC	TFC	ABTS	FRAP	Water	Fructose	Glucose	Remark	State
0	1	itama i	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9	NaN Kedah
1	2	itama ii	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2	NaN Kedah
2	3	itama iii	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4	NaN Kedah
3	4	itama iv	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3	NaN Kedah
4	5	itama v	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1	NaN Kedah

After replacing state codes with state names

REMOVE VALUES



Removing species iteration to have only species name of sample

IMPUTE MISSING VALUES

• • •

```
honey = honey.replace("OOR", "0.0")
```

• • •

```
honey["Water"] = honey["Water"].replace(0.0, 32.00)
```

Impute missing values with a suitable scientific value

CHANGE DATA TYPE

• • •

```
convert_dict = {"Water" : float}  
honey = honey.astype(convert_dict)
```

Convert to suitable data type for model development to have all numerical values

REMOVE UNUSED COLUMNS

```
...  
honey=honey.drop(columns={"Remark", "Sample_ID", "State"})
```

	Species	PH	TPC	TFC	ABTS	FRAP	Water	Fructose	Glucose
0	itama	3.1	147.11	40.43	12.43	7.62	29.1	71.2	70.9
1	itama	3.1	158.81	43.36	15.31	11.41	29.8	70.5	70.2
2	itama	3.0	152.83	44.57	15.21	14.63	30.5	69.7	69.4
3	itama	3.2	203.06	43.97	20.40	20.21	28.8	71.5	71.3
4	itama	3.3	178.09	44.59	17.50	16.19	26.1	74.3	74.1

After removing unnecessary columns

STANDARDIZE VALUES & RANDOMIZE DATASET

```
...  
honey.Species = honey.Species.str.capitalize()
```

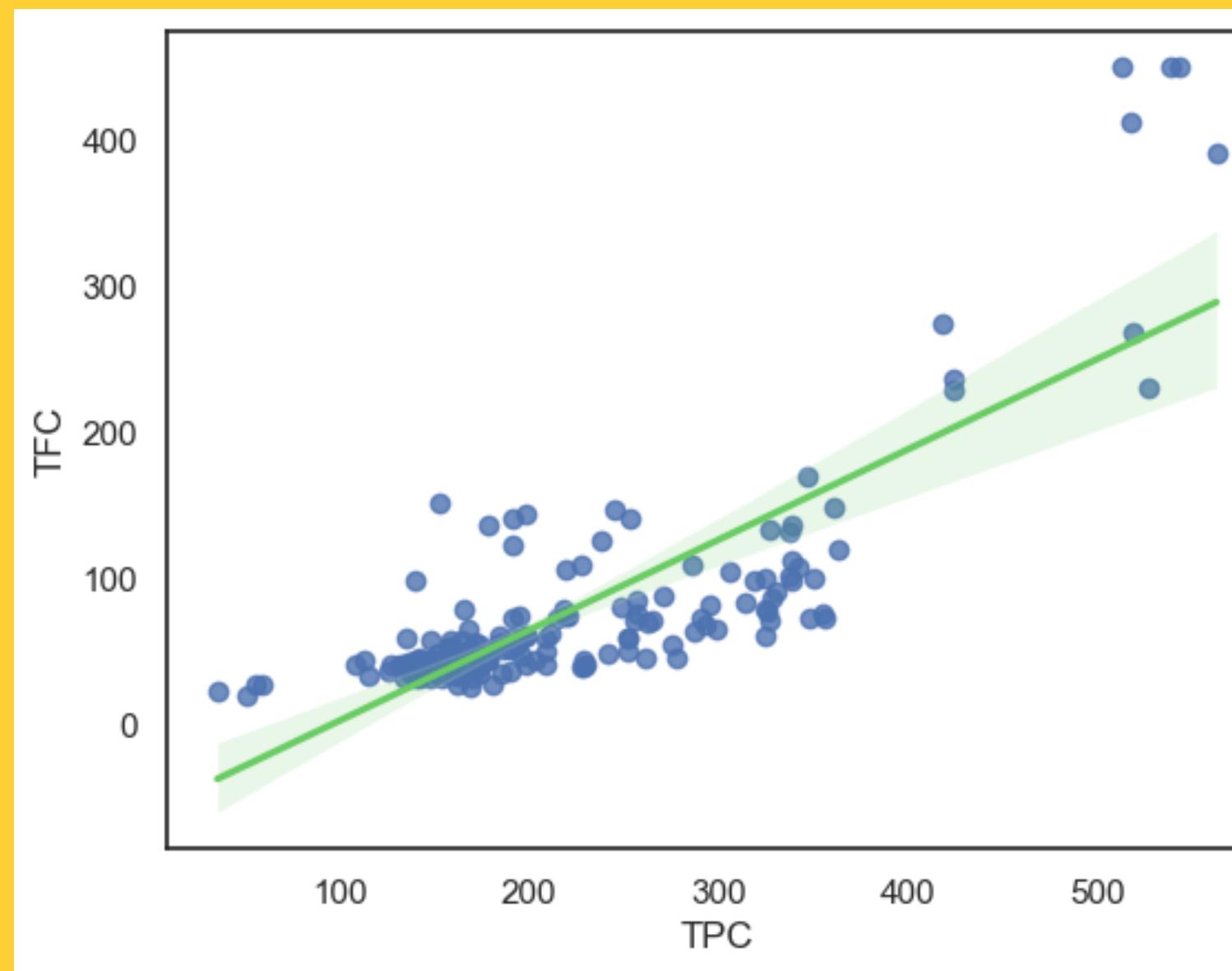
```
...  
honey = honey.sample(frac=1)
```

Species
itama
itama
itama

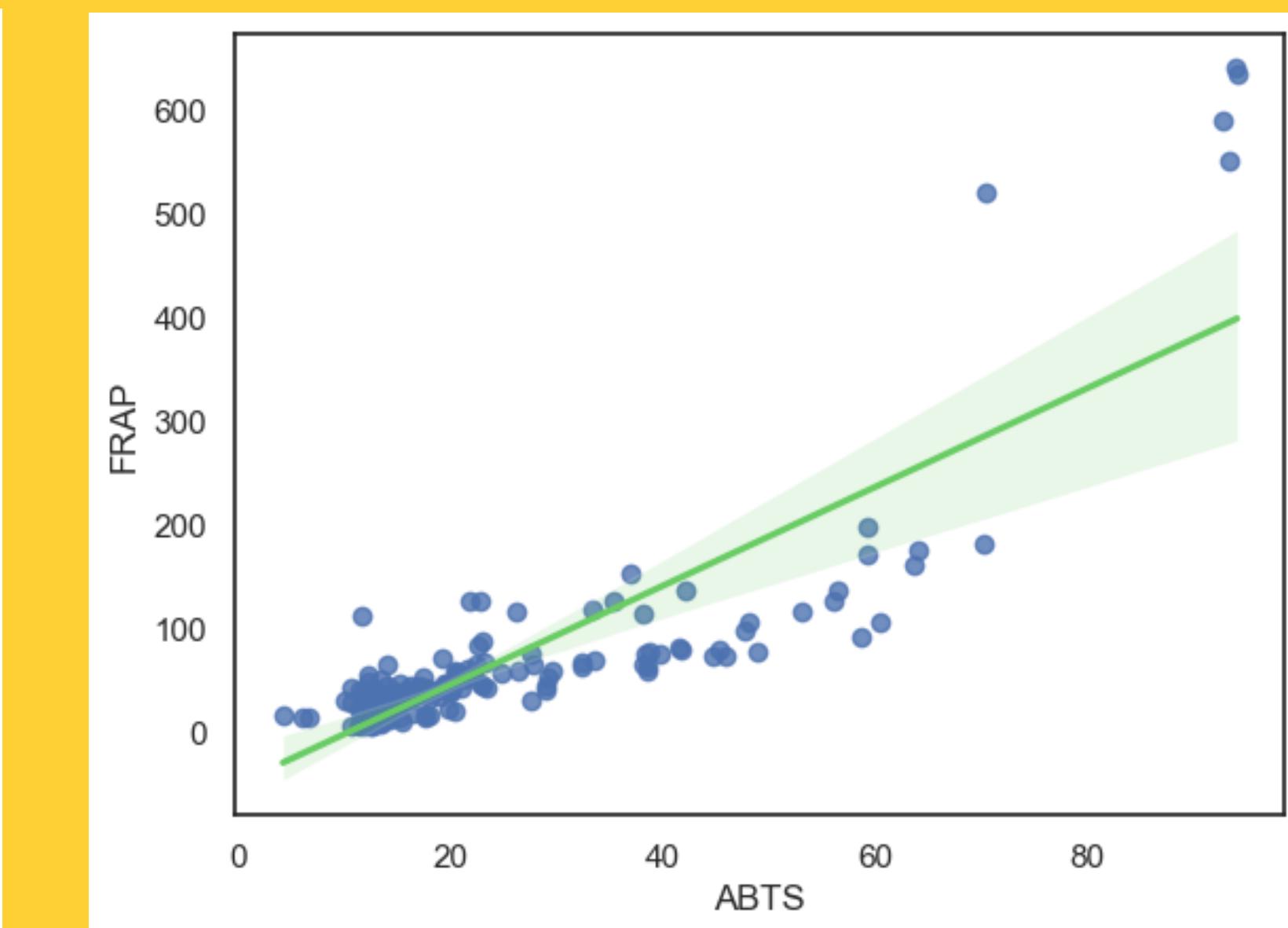
Species
Itama
Thoracica
Itama

DESCRIPTIVE ANALYTICS

Objective 1 : To identify the correlation between different chemical compositions with different antioxidant activities.



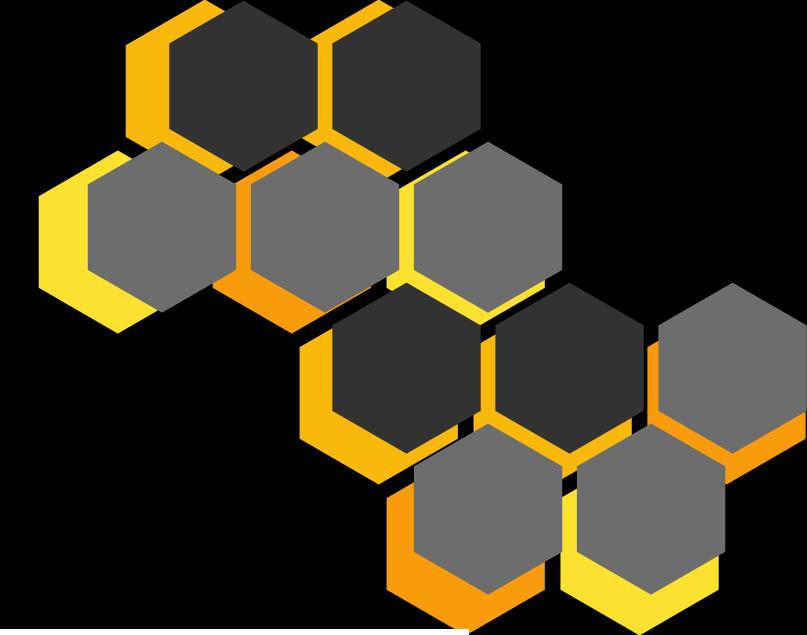
TPC vs TFC
(chemical compounds)



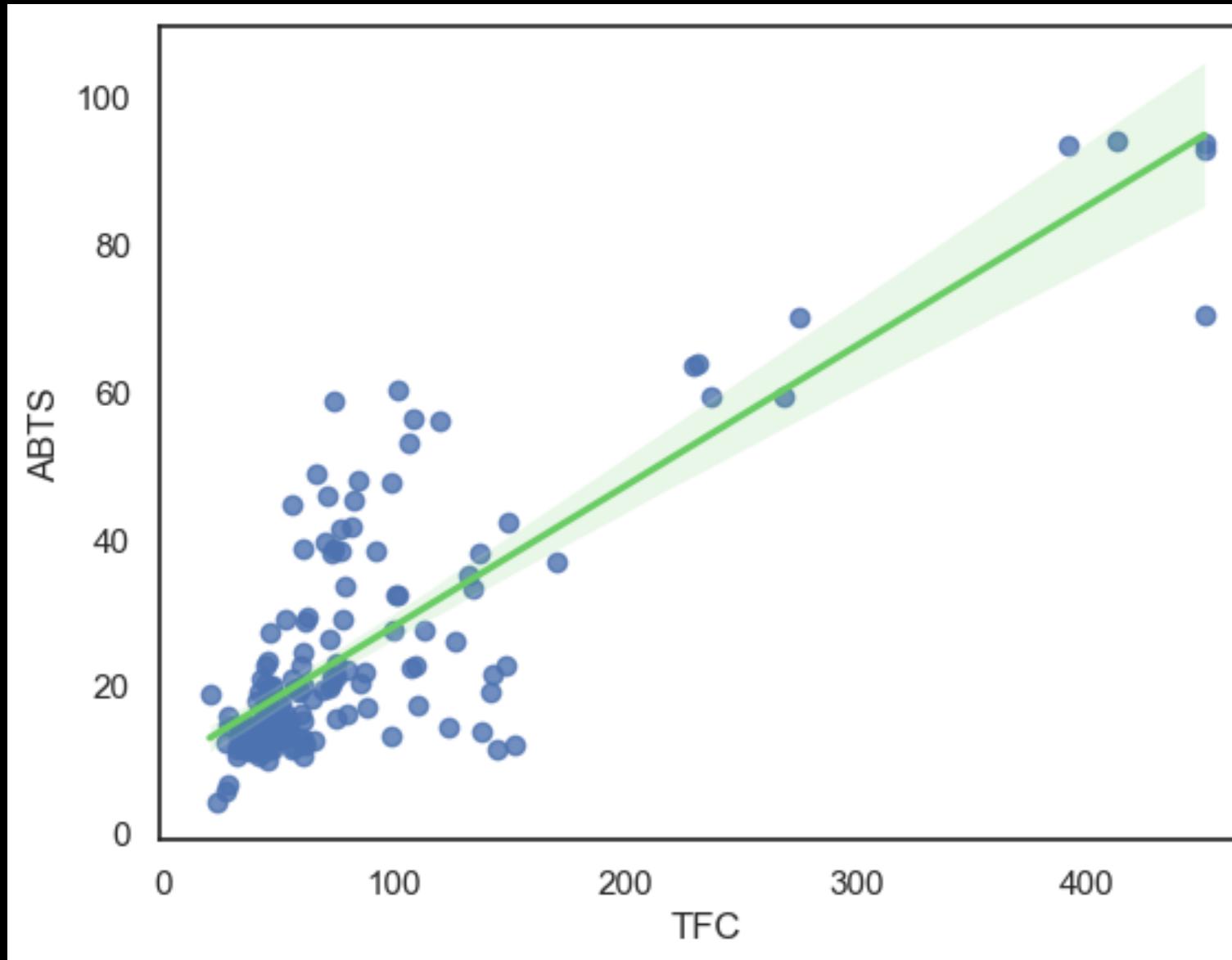
ABTS vs FRAP
(antioxidant activities)

DESCRIPTIVE ANALYTICS

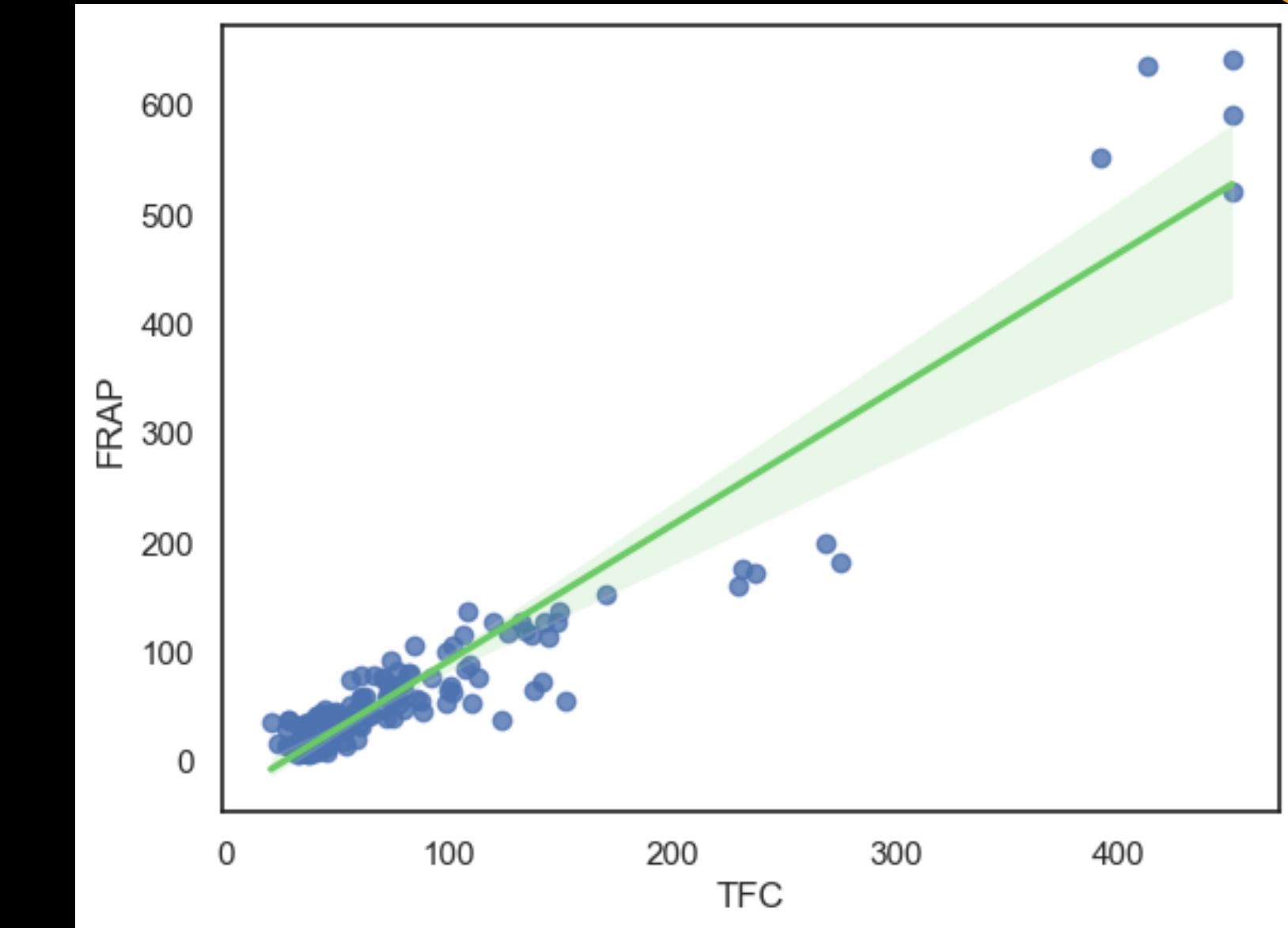
Objective 1 : To identify the correlation between different chemical compositions with different antioxidant activities.



Total Flavonoid Content vs antioxidant assays



TFC vs ABTS



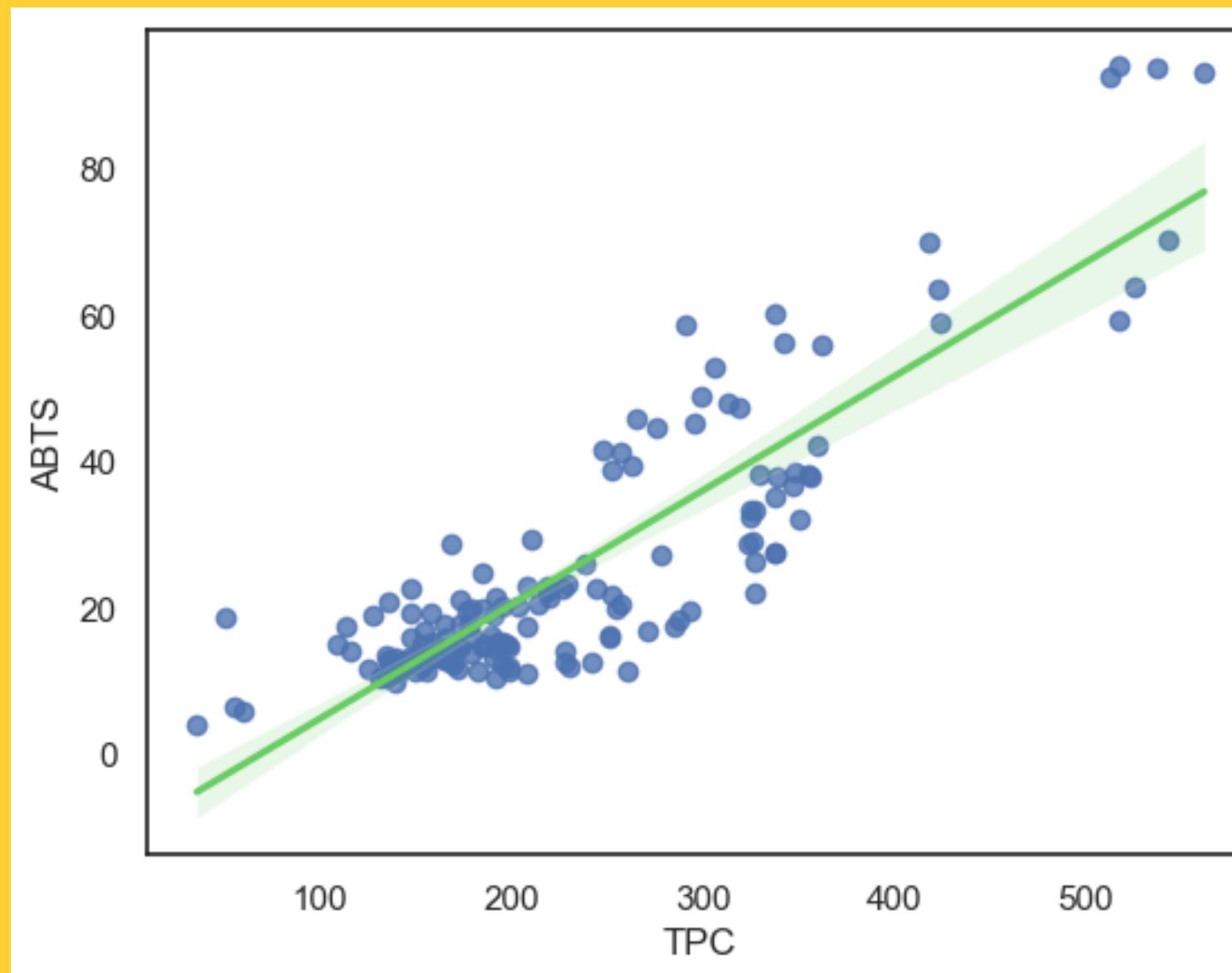
TFC vs FRAP

Linear relationship between TFC and antioxidant assays

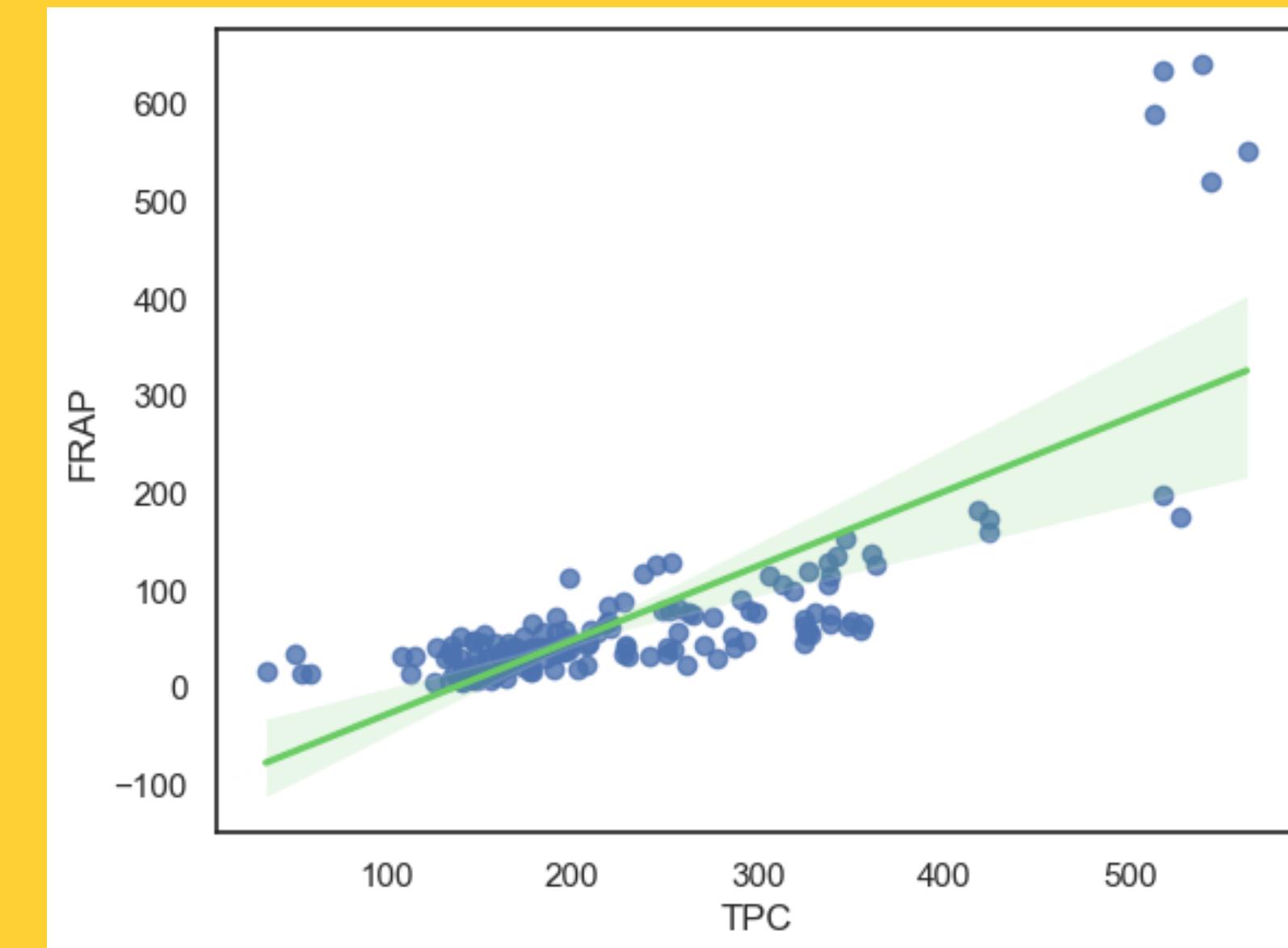
DESCRIPTIVE ANALYTICS

Objective 1 : To identify the correlation between different chemical compositions with different antioxidant activities.

Total Phenolic Content vs antioxidant assays



TPC vs ABTS



TPC vs FRAP

Linear relationship between TPC and antioxidant assays



PREDICTIVE MODELLING

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

Classification Models

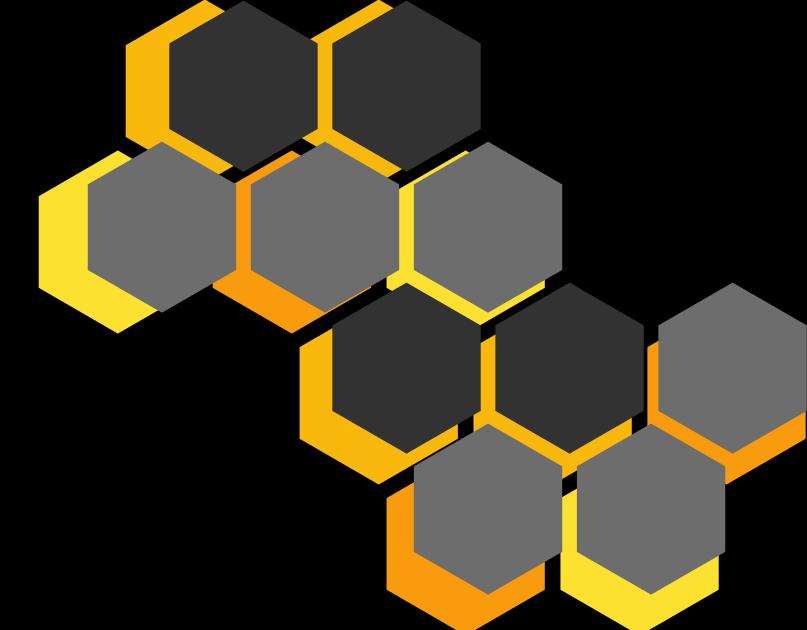
Decision
Trees

Support
Vector
Machine

Random
Forest

Logistic
Regression

K-Nearest
Neighbours



PREDICTIVE MODELLING

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

```
...  
X = honey.iloc[:, 1:]  
y = honey.iloc[:, 0]  
  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)
```

Split dataset into train set and test set with the ratio of 80:20

PREDICTIVE MODELLING

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

```
from sklearn.linear_model import LogisticRegression  
  
classifier = LogisticRegression()  
classifier.fit(X_train, y_train)  
  
# predict  
y_pred = classifier.predict(X_test)
```

Logistic Regression

```
[[1 1 1 0 2 0 0]  
 [1 3 0 0 1 0 0]  
 [0 0 1 1 0 0 0]  
 [0 0 0 1 0 0 0]  
 [2 0 0 1 7 1 3]  
 [0 0 0 0 0 0 0]  
 [1 0 0 0 2 0 2]]  
accuracy is 0.46875
```

```
from sklearn.neighbors import KNeighborsClassifier  
  
classifier = KNeighborsClassifier(n_neighbors=11)  
classifier.fit(X_train, y_train)  
  
# predict  
y_pred = classifier.predict(X_test)
```

K-Nearest Neighbors

```
[[1 0 1 0 1 1 1]  
 [1 3 0 0 0 0 1]  
 [0 0 1 0 0 0 1]  
 [0 1 0 0 0 0 0]  
 [1 1 0 0 6 1 5]  
 [0 0 0 0 0 0 0]  
 [0 0 0 0 3 0 2]]  
accuracy is 0.40625
```

PREDICTIVE MODELLING

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

```
...  
from sklearn.svm import SVC  
  
classifier = SVC()  
classifier.fit(X_train, y_train)  
  
# Predict  
y_pred = classifier.predict(X_test)
```

Support Vector Machine

```
[[ 1  0  0  0  4  0]  
 [ 0  0  0  0  5  0]  
 [ 0  0  0  0  2  0]  
 [ 0  0  0  0  1  0]  
 [ 0  0  0  0 14  0]  
 [ 0  0  0  0  5  0]]  
accuracy is 0.46875
```

```
...  
from sklearn.tree import DecisionTreeClassifier  
  
classifier = DecisionTreeClassifier()  
classifier.fit(X_train, y_train)  
  
# Predict  
y_pred = classifier.predict(X_test)
```

Decision Trees

```
[[ 3  0  0  0  2  0  0]  
 [ 1  3  0  0  0  0  1]  
 [ 0  0  1  0  1  0  0]  
 [ 0  0  0  1  0  0  0]  
 [ 0  0  0  0 12  1  1]  
 [ 0  0  0  0  0  0  0]  
 [ 1  0  0  0  3  0  1]]  
accuracy is 0.65625
```

PREDICTIVE MODELLING

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

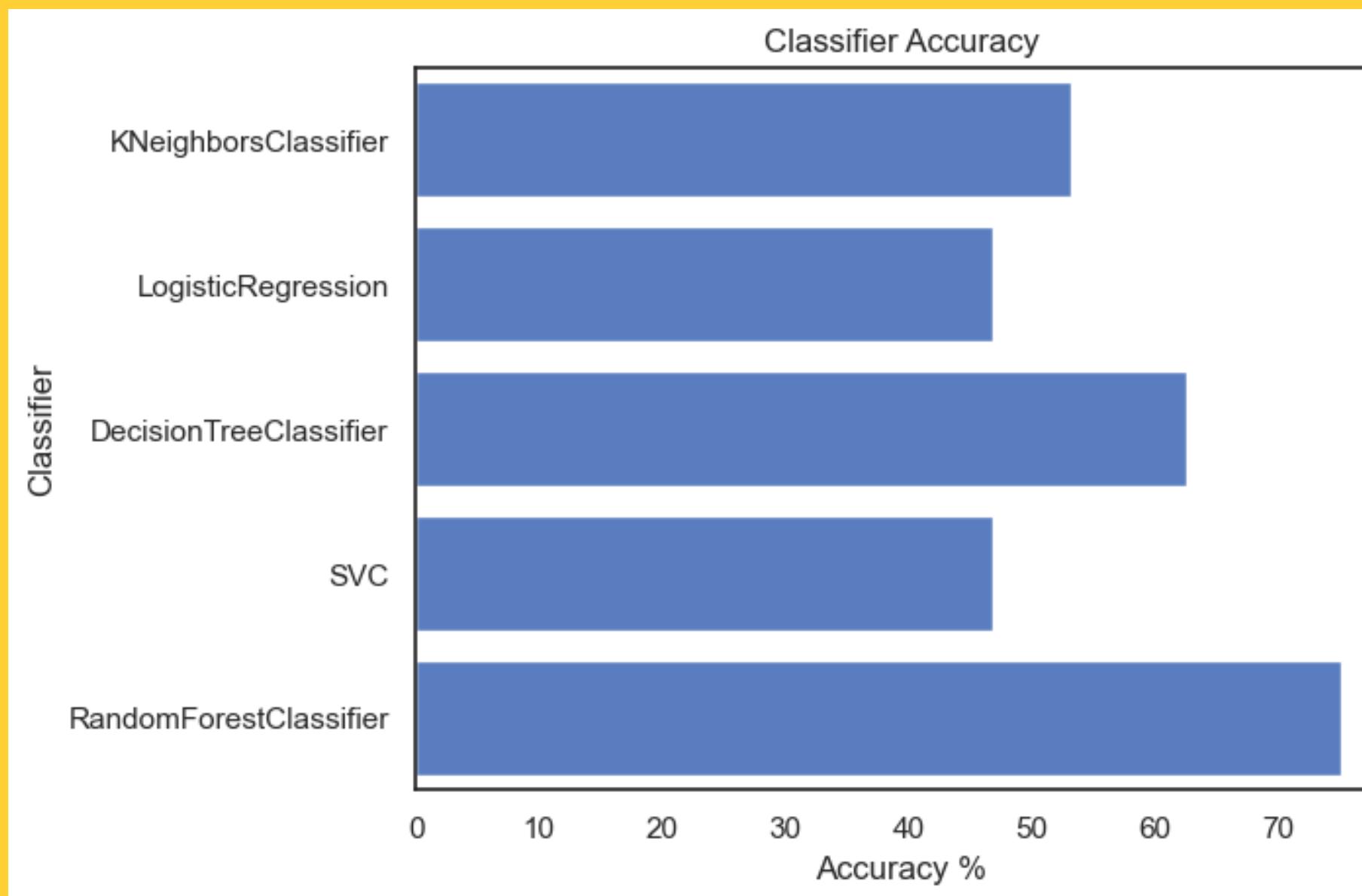
```
...  
from sklearn.ensemble import RandomForestClassifier  
  
classifier = RandomForestClassifier(n_estimators = 100)  
classifier.fit(X_train, y_train)  
  
# Predict  
y_pred = classifier.predict(X_test)
```

Random Forest

```
[[ 4  0  0  0  1  0]  
 [ 0  5  0  0  0  0]  
 [ 0  0  1  0  0  1]  
 [ 0  0  0  1  0  0]  
 [ 2  0  0  0  11  1]  
 [ 0  0  0  0  2  3]]  
accuracy is 0.78125
```

MODEL EVALUATION

Objective 2 : To develop a classification model to classify species of stingless bee based on samples of honey.

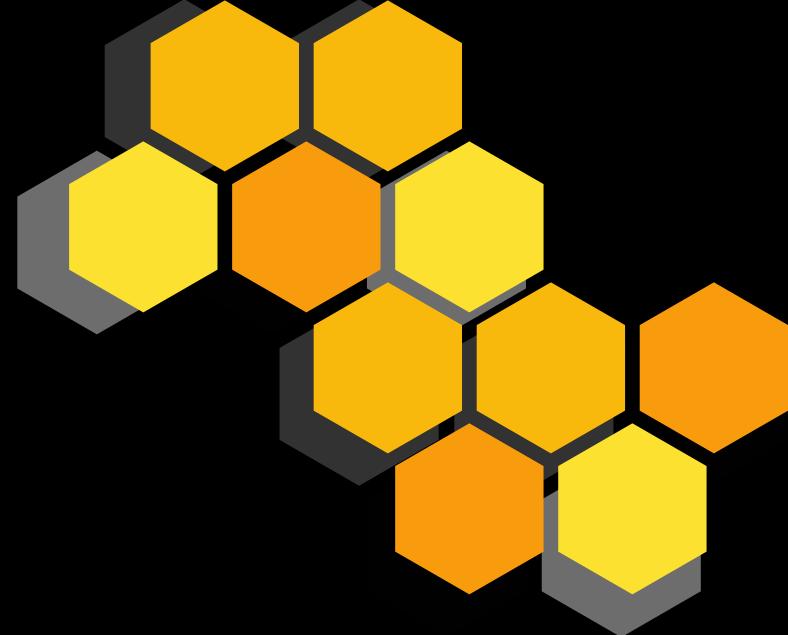


Classifiers accuracy

Random Forest will be chosen as the main model for classification with the accuracy of 78%

MODEL DEPLOYMENT

Objective 3 : To deploy the classification as an interactive application that allows user to predict species of stingless bees based on honey attributes.

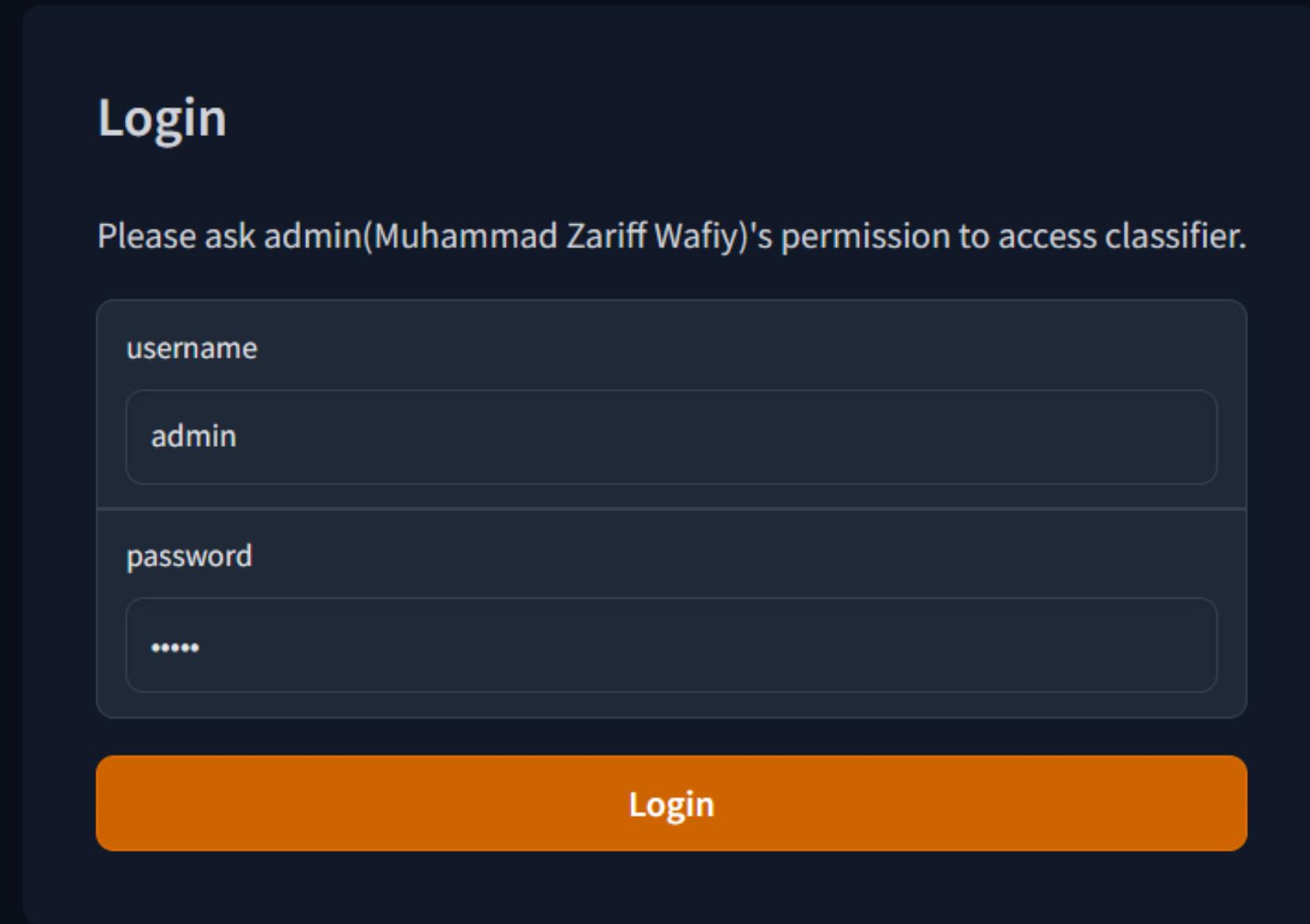


```
...  
  
import pickle  
  
# save the fitted model  
with open("RandomForest.pkl", "wb") as f:  
    pickle.dump(classifier, f)  
  
# load  
with open("RandomForest.pkl", "rb") as f:  
    classifier = pickle.load(f)
```

Save & Load Model

MODEL DEPLOYMENT

Objective 3 : To deploy the classification as an interactive application that allows user to predict species of stingless bees based on honey attributes.



The image shows a login page with a dark background. At the top left, the word "Login" is written in white. Below it, a message in white text reads: "Please ask admin(Muhammad Zariff Wafiy)'s permission to access classifier." There are two input fields: one for "username" containing "admin" and another for "password" containing ".....". At the bottom is a large orange button with the word "Login" in white.

Login page

Stingless Bee Honey Classifier

Utilize this classifier to determine which species of stingless bee is your honey from! Note that if you do not know the value of a specific component/attribute, just leave it as it is. The default value corresponds to the mean of the attribute.

pH value 3.11

Total Phenolic Content, TPC 227.23

Total Flavonoid Content, TFC 82.13

ABTS 24.78

FRAP 69.62

Water Content 27.2

Fructose Content 67.42

Glucose Content 67.25

Does it contain other kinds of sugar e.g cane sugar, corn syrup, palm sugar, etc.?

Clear

MODEL DEPLOYMENT

Objective 3 : To deploy the classification as an interactive application that allows user to predict species of stingless bees based on honey attributes.

Stingless Bee Honey Classifier

Utilize this classifier to determine which species of stingless bee is your honey from! Note that if you do not know the value of a specific component/attribute, just leave it as it is. The default value corresponds to the mean of the attribute.

pH value 3.3

Total Phenolic Content, TPC 139.3

Total Flavonoid Content, TFC 198.3

ABTS 30.65

FRAP 432.43

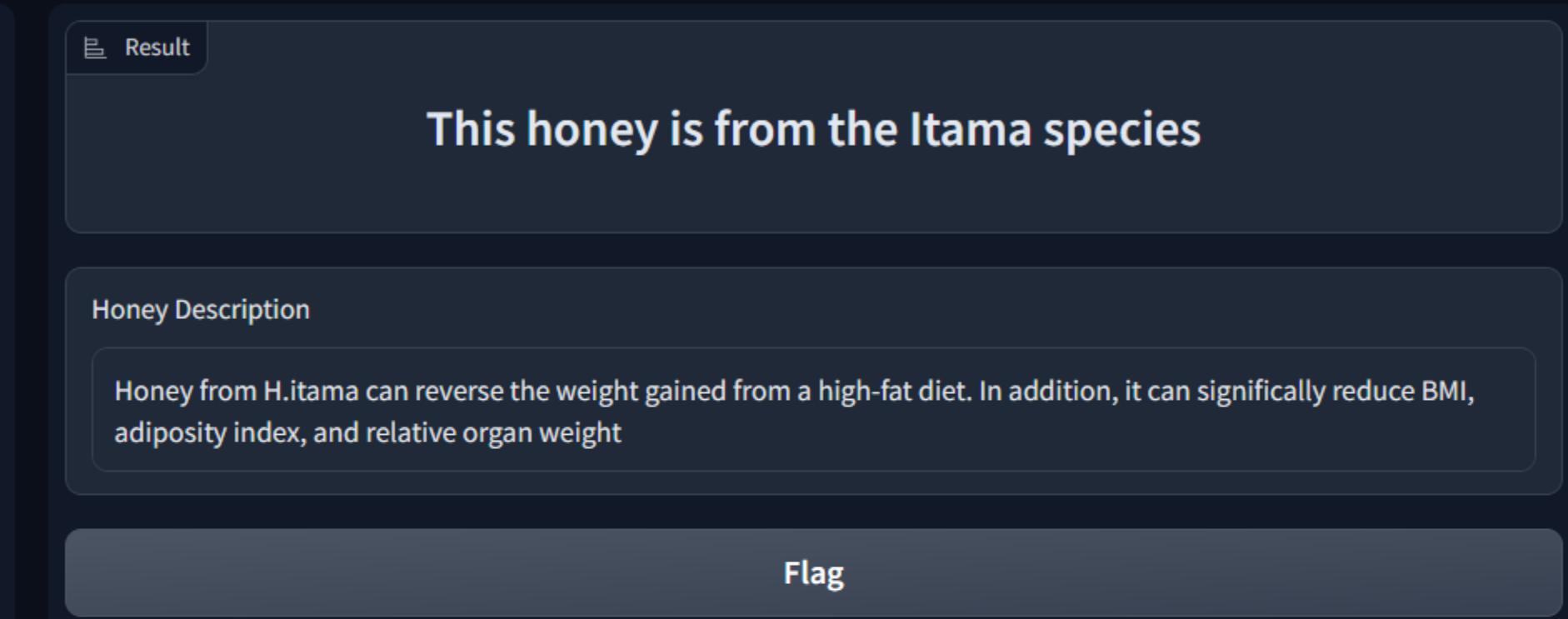
Water Content 29.38

Fructose Content 72.4

Glucose Content 73.7

Does it contain other kinds of sugar e.g cane sugar, corn syrup, palm sugar, etc.?

Clear



MODEL DEPLOYMENT

Objective 3 : To deploy the classification as an interactive application that allows user to predict species of stingless bees based on honey attributes.

REFERENCES

- Abd Jalil, M. A., Kasmuri, A. R., & Hadi H. (2017). Stingless Bee Honey, the natural wound healer: A Review. *Skin Pharmacology and Physiology*, 30(2), 66-75. Retrieved November 5, 2022, from <https://doi.org/10.1159/000458416>
- Fletcher, M.T., Hungerford, N.L., Webber, D. et al. Stingless bee honey, a novel source of trehalulose: a biologically active disaccharide with health benefits. *Sci Rep* 10, 12128 (2020). <https://doi.org/10.1038/s41598-020-68940-0>
- Pimentel, T.C., Rosset, M., Sousa, J.M., Oliveira, L. I., Mafaldo, I.M., Pintado, M. M., Souza, E.L., & Magnani, M. (2021). Stingless Bee Honey: An overview of health benefits and main market challenges. *Journal of Food Biochemistry*, 46(3). <https://doi.org/10.1111/jfbc.13883>
- Razali, M., Zainal, Z., Maulidiani, M., Shaari, K., Zamri, Z., Mohd Idrus, M., Khatib, A., Abas, F., Ling, Y., Rui, L., & Ismail, I. (2018). Classification of raw stingless bee honeys by bee species origins using the NMR-and LC-MS-based metabolomics approach. *Molecules*, 23(9), 2160. <https://doi.org/10.3390/molecules23092160>
- Souza, E.C.A., Menezes, C. & Flach, A. Stingless bee honey (Hymenoptera, Apidae, Meliponini): a review of quality control, chemical profile, and biological potential. *Apidologie* 52, 113–132 (2021). <https://doi.org/10.1007/s13592-020-00802-0>



**THANK
YOU**