D.Y.Patil College of Engineering Akurdi.

A
Project Report
On

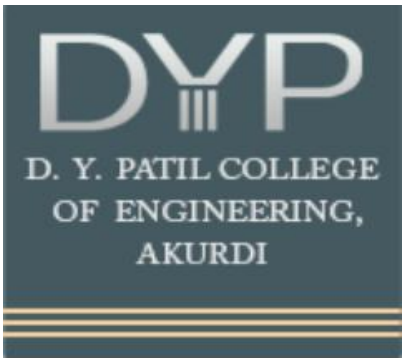
Mushroom Classification

SUBMITTED BY

NAME: Vikas Mane(B1921152)
NAME: Abhijeet Bhalekar(B1921155)

CLASS: BE-B

GUIDED BY
Prof. P.B.Deshmukh



COMPUTER ENGINEERING DEPARTMENT

Academic Year:2021-22

D.Y.Patil College of Engineering,Akurdi.

CERTIFICATE

This is to certify that **Mr. Vikas Mane(B1921152),Abhijeet Bhalekar(B1921155)**, a students of B.E. (Computer Engineering Department)
Batch 2021-2022, has satisfactorily completed a project report on "**Mushroom Classification**" under the guidance of **Prof. P.B.Deshmukh**. towards the partial fulfillment of the fourth year Computer Engineering Semester I of Pune University.

Prof. P.B.Deshmukh	Dr. D.A.Potey
Project Guide	Head of Department

CONTENTS

1. PROBLEM STATEMENT

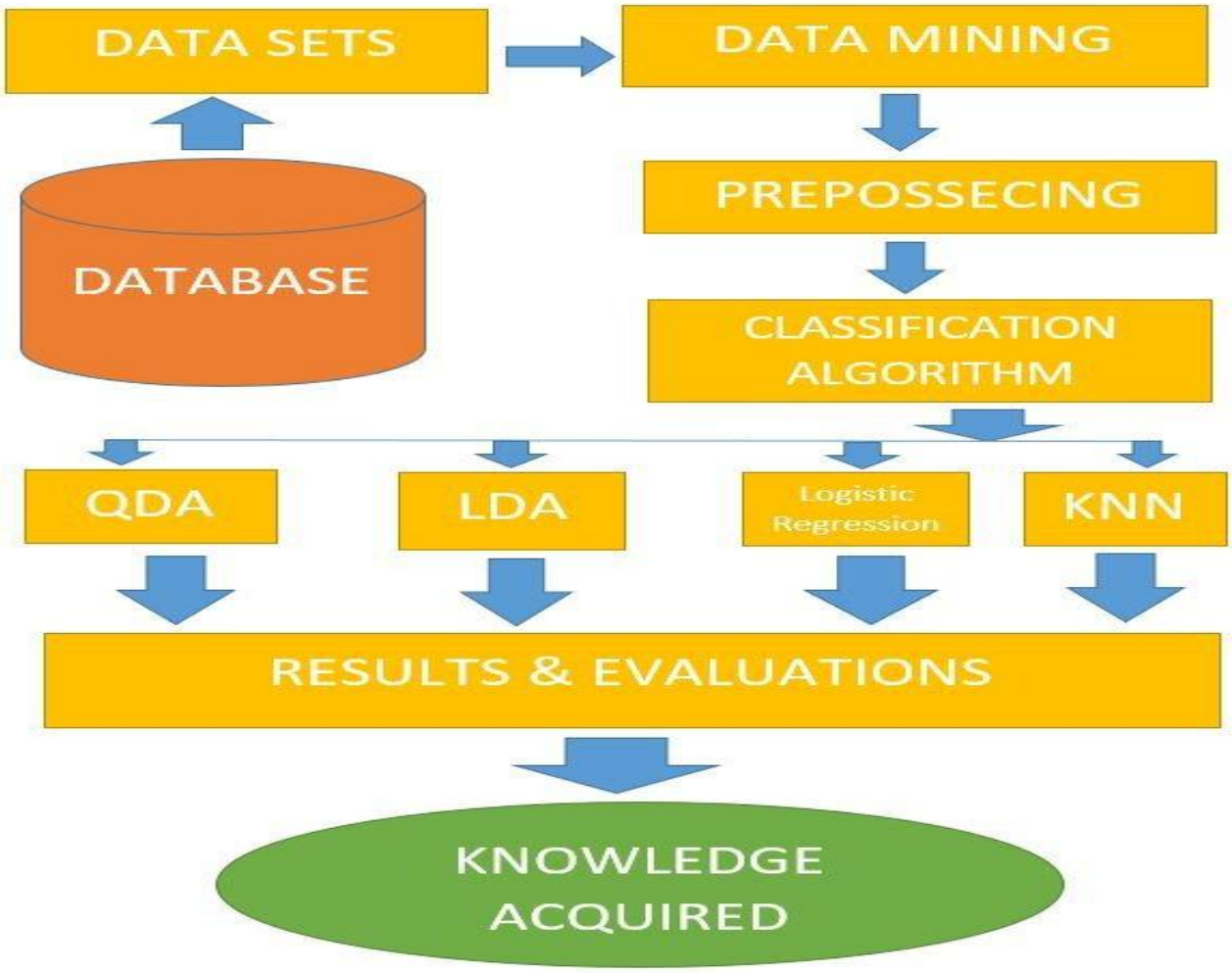
- Classifying the categories of mushroom using data mining techniques

2. INTRODUCTION

- Data Mining is a process of extracting useful information from data warehouses or from bulk data. This article contains the Most Popular and Frequently Asked Interview Questions of Data Mining along with their detailed answers. These will help you to crack any interview for a data scientist job. So let's get started.
- The following activities are carried out during data mining:
    - ✓ Classification
    - ✓ Clustering
    - ✓ Association Rule Discovery
    - ✓ Sequential Pattern Discovery
    - ✓ Regression
    - ✓ Deviation Detection

- In this project we are classifying all the types of mushrooms though the data mining techniques

3.METHODOLOGY



4.OUTPUT

1.Mushroom data analysis

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
```

2.First five rows of dataset

Out[5]:

	0	1	2	3	4
class	p	e	e	p	e
cap-shape	x	x	b	x	x
cap-surface	s	s	s	y	s
cap-color	n	y	w	w	g
bruises	t	t	t	t	f
odor	p	a	l	p	n
gill-attachment	f	f	f	f	f
gill-spacing	c	c	c	c	w
gill-size	n	b	b	n	b
gill-color	k	k	n	n	k
stalk-shape	e	e	e	e	t
stalk-root	e	c	c	e	e
stalk-surface-above-ring	s	s	s	s	s
stalk-surface-below-ring	s	s	s	s	s
stalk-color-above-ring	w	w	w	w	w
stalk-color-below-ring	w	w	w	w	w
veil-type	p	p	p	p	p
veil-color	w	w	w	w	w
ring-number	o	o	o	o	o
ring-type	p	p	p	p	e
spore-print-color	k	n	n	k	n
population	s	n	n	s	a

Out[15]:

Predictions	0	1
Actual		
0	1257	0
1	3	1178

Out[17]:

Predictions	0	1
Actual		
0	2951	0
1	5	2730

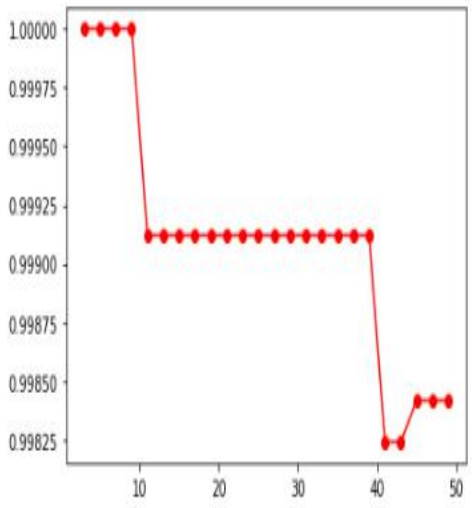
Output

3. Checking for null values

```
Out[6]: class                       0
        cap-shape                   0
        cap-surface                 0
        cap-color                   0
        bruises                     0
        odor                        0
        gill-attachment             0
        gill-spacing                0
        gill-size                   0
        gill-color                  0
        stalk-shape                 0
        stalk-root                  0
        stalk-surface-above-ring    0
        stalk-surface-below-ring    0
        stalk-color-above-ring      0
        stalk-color-below-ring      0
        veil-type                   0
        veil-color                  0
        ring-number                 0
        ring-type                   0
        spore-print-color           0
        population                  0
        habitat                     0
        dtype: int64
```

4. Accuracy graph

```
Out[18]: [<matplotlib.lines.Line2D at 0x29d8fa3a7f0>]
```



6.REFERENCES

I. R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules (1994) Proc. 20th Int. Conf. Very Large Data Bases, VLDB-94.

II. Bayardo, R. and R. Srikant, Technological Solutions for Protecting Privacy, IEEE Computer, Sep 2003

III. Chernoff, H. (1973). Using faces to represent points in k-dimensional space graphically. Journal of American Statistical Association, 68, 361-368.

IV. https://www.ibm.com › Cloud › Cloud Learn

V. https://en.wikipedia.org › wiki › Data_mining

VI. https://www.sas.com/en_us/insights/analytics/data-mining.html