VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590 018, KARNATAKA

#### vtulogo

*A Mini Project Report on*

“Pizza or Pasta classifier using KNN algorithm”

Submitted in the partial fulfillment for the requirements for the Data Minning with Mini Project .

in

#### **INFORMATION SCIENCE AND ENGINEERING**

#### By

**Mr. Achutha Sandesh S Chatra USN:1BY18IS003**

**Ms.Chaitanya M USN: 1BY18IS037**

**Mr. Mithun G USN: 1BY18IS066**

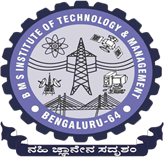
**Mr. Raghavendra K M USN:1BY18IS093**

Under the guidance of

**Mr.Chandrashekar**

Assistant Professor

Department of ISE, BMSIT&M.

****

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

# BMS INSTITUTE OF TECHNOLOGY & MANAGEMNT

**YELAHANKA, BENGALURU-560064**

**2018-2022**

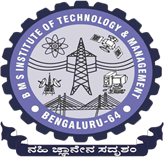
**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

**BELAGAVI – 590 018, KARNATAKA**

**BMS INSTITUTE OF TECHNOLOGY & MANAGEMENT**

**YELAHANKA, BENGALURU-560064**

**DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING**

****

**CERTIFICATE**

This is to certify that the Project work entitled **“Pizza or Pasta classifier using KNN algorithm”** is a bonafide work carried out by **Mr. Achutha Sandesh S Chatra(1BY18IS003)**, **Ms.Chaitanya M (1BY18IS037), Mr.Mithun G(1BY18IS066), Mr.Raghavendra K M (1BY18IS093)** in partial fulfillment of Data Minning with Mini Project for the award of **Bachelor of Engineering Degree in Information Science and Engineering** of the Visvesvaraya Technological University, Belagavi during the year 2018-2022. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in this report. The project report has been approved as it satisfies the academic requirements in respect of Mini Project work for the B.EDegree.

**Signature of the Guide Signature of the HOD**

Mr.Chandrashekar Dr.Pushpa k

Assistant Professor Professor and Head

Department of ISE Department of ISE

**ACKNOWLEDGEMENT**

We are happy to present this mini project after completing it successfully. This mini project would not have been possible without the guidance, assistance and suggestions of many individuals. We would like to express our deep sense of gratitude and indebtedness to each and every one who has helped us make this mini project a success.

We heartily thank our **Principal, Dr. Mohan Babu G.N,** **B M S Institute of Technology & Management** for his constant encouragement and inspiration in taking up this mini project.

We heartily thank our **Head of Department Dr. Pushpa k, Dept. of Information Science** **and** **Engineering, B M S Institute of Technology& Management** for his constant encouragement and inspiration in taking up this mini project.

We gracefully thank our Project guide**, Mrs. Chandrashekar, Asst. Professor, Dept. of** **Information Science and Engineering,** for her encouragement and advice throughout the course of the mini project work.

Special thanks to all the staff members of Information Science Department for their help and kind co-operation.

We also thank our parents and friends for their unconditional love and encouragement and support given to us in order to finish this precious work.

Last but not the least we would like to thank God for giving us the strength and motivation through the course of this Project.

**ABSTRACT**

Pizza or pasta classifier is a machine learning model that uses the k-nearest neighbors (KNN) algorithm to classify whether a given image contains a pizza or pasta. The data set is obtained from bing-image-loader which downloads the specified number of images when passed (size of dataset) as parameter. cv2. imread() method loads an image from the specified file. cvtColor() method is used to convert an image from one color space to another. The matplotlib function imshow() creates an image from a 2-dimensional numpy array. The image will have one square for each element of the array. The color of each square is determined by the value of the corresponding array element and the color map used by imshow() . shape() is used to get the dimensions of the image in the format of (height,width,channels). resize() is used to change the dimensions of an image, be it width alone, height alone or changing both of them. flatten() is used to convert Multi-Dimensional to 1-D arrays to conserve memory.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| CHAPTER | TOPIC |
| 1. | Introduction |
| 2. | * 1. Implementation   2. Modules used   3. Working |
| 3. | * 1. Applications |
| 4. | * 1. Conclusion |
| 5. | * 1. References |

**INTRODUCTION:**

It is a supervised machine learning algorithm that can be used to solve both classification and regression problems. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification).

**K-NN ALGORITHM**

* K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
* K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
* K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
* K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
* K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
* It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
* KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
* **Example:** Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

****

## **Why do we need a K-NN Algorithm?**

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram: 

# IMPLEMENTATION

# The model is trained and implemented as follows.

### Modules used:

#### **bing-image-loader:** Python library to download bulk of images from Bing.com. This package uses an async url, which makes it very fast while downloading.

#### **matplotlib.pyplot:** pyplot is mainly intended for interactive plots and simple cases of programmatic plot generation

#### **sklearn: os:** This module provides a portable way of using operating system dependent functionality. If you just want to read or write a file see [open()](https://docs.python.org/3/library/functions.html#open), if you want to manipulate paths, see the [os.path](https://docs.python.org/3/library/os.path.html#module-os.path) module

#### **skimage.io:** Utilities to read and write images in various formats. Concatenate all images in the image collection into an array.

#### **skimage.transform:** We can resize images with the skimage. transform. resize() function.

#### **numpy**: NumPy is the fundamental package for scientific computing in Python. NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data. Typically, such operations are executed more efficiently and with less code than is possible using Python's built-in sequences

#### **pandas:** Pandas is mainly used for data analysis. Pandas allows importing data from various file formats such as comma-separated values, JSON, SQL, Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features.

#### The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

#### **Working:**

* The data set is obtained from bing-image-loader which downloads the specified number of images when passed (size of dataset) as parameter.
* cv2. imread() method loads an image from the specified file. cvtColor() method is used to convert an image from one color space to another.
* The matplotlib function imshow() creates an image from a 2-dimensional numpy array. The image will have one square for each element of the array. The color of each square is determined by the value of the corresponding array element and the color map used by imshow() .
* shape() is used to get the dimensions of the image in the format of (height,width,channels).
* resize() is used to change the dimensions of an image, be it width alone, height alone or changing both of them.
* flatten() is used to convert Multi-Dimensional to 1-D arrays to conserve memory.
* os.path.split(*)* method in Python is used to Split the path name into a pair *head* and *tail*. Here, *tail* is the last path name component and *head* is everything leading up to that.
* Why flat\_data.shape() returns (40,7500) because 40 represents 40 images and 7500 is 2500 pixels \* 3 color(RGB)

**APPLICATIONS**

* Can be improved and made to suggest nutritional value of classified image so that you won’t mess up your diet.
* Can be implemented for other confusing diets like burger or sandwich, tea or coffee etc.
* Can be implemented in websites to give captions for the images
* Can be used by the students to identify or learn about the given input image/object.

**CONCLUSION**

* The project uses data collected from the bing search to train the machine learning model.
* The model is effective in recognising a given input image as Pizza or Pasta and can be used for similar recognising tasks.
* Efficiency of 96.2% is achieved from the current trained model.

**REFERENCES:**

An excellent website to understand KNN and its beginner friendly.

* + <https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e>
  + <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

A concrete guide for image processing.

<https://towardsdatascience.com/image-processing-opencv-vs-pil-a26e9923cdf3>

How to apply KNN to classify images.

<https://www.pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/>