**PYTHON FOR MACHINE LEARNING**

**TITLE OF THE PROJECT**

**LETTER RECOGNITION**

**Submitted by** Group E3

**Sai Abhishek - BL.EN.U4AIE21015**

**Ruchith Balaji - BL.EN.U4AIE21017**

**Chillakuru Hari - BL.EN.U4AIE21038**

**BACHELOR OF TECHNOLOGY**

**IN**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**



**AMRITA SCHOOL OF ENGINEERING, BANGALORE**

**AMRITA VISWA VIDYAPEETHAM**

**BANGALORE- 560035**

**Literature Survey:**

Digitization of machine printed or Letter documents have become very popular with the advancements in computing and technology. Humans have tried to automatized their work by replacing themselves with machines. The transformation from manual to automatization gave rise to several research areas and text recognition is one among them.

<https://ieeexplore.ieee.org/document/9167649>

Deep learning and machine learning techniques have been proved to be very suitable for optical character recognition. In this work, an up-to-date overview of four machine learning and deep learning architectures, viz., Support vector machine, Artificial neural network, Naive Bayes and Convolutional neural network have been discussed in detail

<https://link.springer.com/article/10.1007/BF00114162>

<https://www.sciencedirect.com/science/article/abs/pii/S0364021385800124>

**ABSTRACT:**

Letter recognition has been an active and challenging area. Letter

recognition system plays a very important role in today’s world. At present time

it is very difficult to find the correct meaning of Lettered documents as

different people have different ways of writing digits or texts.

There are many areas where we need to recognize the words, alphabets and

digits like postal addresses, bank cheques etc.

The main aim of this project is to predict the Alphabet corresponding to the

Attributes of the letter which are given in the Data set.

Output of the project that allows users to predict the digits drawn on the interface using machine learning and shows the results in percentage how likely the prediction is correct. The project will be built using python and the application will be trained with the help of dataset from Kaggle using K-Nearest Neighbour, Support Vector Machines, Decision trees. Upon all which has the Highest Accuracy, then that Algorithm is preferred to predict the Letters.

**Purpose of Study:**

The purpose for this project is to allow us to explain the implementation of the Machine learning models and how the models use the dataset to analyse and predict the outcome.

This project allows us to study the K-Nearest Neighbour, Support Vector Machines, Decision trees of the Letter Recognizer and the working of the model to see how the Machine Learning process works in the background. It shows the working of how the number fed into the model is recognized and displays the number that is recognized after analysis and comparing it with the dataset fed to the model. With the help of the Letter Recognition dataset and the K-Nearest Neighbour, Support Vector Machines, Decision trees. The project can explain the purpose, working and the benefits of Machine learning in our daily lives

**Data Set:**

To make this model, the data set needs to be imported which is very important in the analysis and recognizing the digits which are fed into the model. The dataset used in the project is called the Letter Recognition dataset.

The Letter Recognition is a large dataset of Letters with different attributes like size of letter, height of the letter, how much space needed to fit, Width of the letter, space needed for edge, etc Machine Learning models for training various Recognition systems. This dataset contains 20,000 letters and each letter has its own attributes as mentioned above.

**Making a Model:**

**Training Data**

**Training for ML Algorithm**

**Model Input**

**Input Data**

**ML Algorithm**

**Prediction**

**Successful Model**

1. **K-Nearest Neighbour:**

* From the given Data set find dependent and independent Variables.
* Divide the variables to training and testing data, it means Data pre-processing.
* Fit the train and test data.
* Using the KNN Classifier taking n neighbours find the predicted for test data
* From predicted data find Accuracy, Precision, Recall, F1-Score etc.

1. **Support Vector Machines:**

* From the given Data set find dependent and independent Variables.
* Divide the variables to training and testing data, it means Data pre-processing.
* Fit the train and test data.
* Using the SVC Classifier taking path as linear, curve find the predicted for test data
* From predicted data find Accuracy, Precision, Recall, F1-Score etc.

1. **Decision Trees:**

* From the given Data set find dependent and independent Variables.
* Divide the variables to training and testing data, it means Data pre-processing.
* Fit the train and test data.
* Using the Decision Tree Classifier taking Criterion as entropy find the predicted for test data
* From predicted data find Accuracy, Precision, Recall, F1-Score etc.

Among all the three models, the model having more accuracy is preferred.

**CONCLUSION:**

In this project, the Letter Recognizer has been implemented and is able to recognize the Letters of different Sizes. Among the three K-Nearest Neighbour, Support Vector Machines, Decision trees; KNN is one of the most widely used machine learning algorithms which has been trained and tested on the given dataset in order to compare and analyse. With this Machine learning technique, a high amount of accuracy can be obtained.