# SENTIMENT ANALYSIS USING COMPUTER VISION

Prof. **B. Prajna** Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh
**Lanka Adi Lakshmi, Tadi Aryani, B.V.S.Kedareswari, Chenna Sravya** Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh
*prajna.mail@gmail.com*

## ABSTRACT

Enhancing modern day computers or systems to recognize various sentimental analysis by using the facial expressions and to understand human emotions from them in real time is an emerging research subject. Through this paper, we put forward a solution to recognize emotions by understanding different facial expressions by collecting live video through a Open CV application. By the use of open cv we need to video stream live feed captured through the local camera or web camera attached to the system or computer. The video captured is fed to various image extraction techniques. We have declared a dictionary which contains the emotions in the dataset and a list with the emotions which we want to observe.We took the Image datasets as inputs and We considered 75% of data for training and 25% of data for testing.Initialized a classifier model and fitted it. we found accuracy for 10 different classifiers and picked one classifier with more accuracy. By using that classifier, we found the count of files of each ofthe emotions in each dataset. Step 11: We drew a classification report for each of the datasets. We drew a confusion matrix for each of the datasets. The paper is concluded by discussing the outcomes of our project and ways to improve the efficiency of the model. The scope of the project is also analyzed to enhance technologies developed in the near future.

**Keywords :** *Machine Learning ,Long Short-Term Memory, Recurrent Neural Networks , Logistic Regression Algorithm,Linear Regression Algorithm, AprioriAlgorithm, .K Means Clustering Algorithm.*

## 1. INTRODUCTION

Predicting the action of features is a kind of applications from human-robot collaboration and autonomous robot navigation to exploring abnormal situations in surveillance videos and activity-aware service algorithms for private or health care purposes .As an example,in autonomous health care services, consider an agent- monitoring a patient's activities, trying to predict if the patient is losing her/his balance.If the agent is capable of predicting the following action, it could determine whether s/he might fall and take an action to try to stop it.

This is result of monitoring a 26-year-old man in a three-room apartment where 14 binary sensors were installed. These sensors were installed in locations like doors, cupboards, refrigerators, freezers or toilets. Sensor data for 28 days is collected for a complete of 2120 sensor events and 245 activity instances. The annotated activities were the following: "Leave House", "Use a Toilet", "Take a Shower", "Go To Bed", "Prepare Breakfast", "Prepare a Dinner" and "Get a Drink". during this specific case, the sensors were mapped one to at least one to actions, leading to the subsequent set of actions: "Use Dishwasher","Open a Pans Cupboard", "Toilet a Flush", "Use a Hall Bedroom Door", "Open a Plates Cupboard", "Open Cups Cupboard", "Open Fridge",Use Microwave", "Use Hall Bathroom Door", "Use Washing machine", "Use Hall Toilet Door", "Open Freezer", "Open Groceries Cupboard" and "Use Front door".

Post this the sensors have been started training In order to make the training process more streamlined, we apply the sensor to action mappings offline this enables us to coach the deep neural model faster while still having the raw sensor data because the input to try and do the training, we use actions because the input to predict the subsequent action. That is the training examples are the sequences of actions, and also the label is that the next action which follows that sequence, being a supervised learning problem.


Copyright @ 2021 Authors

## 2. Libraries Modules and Features used:

### 2.1 Libraries:

**Numpy:** NumPy is a Python package that is used to perform scientific computations. It is a library that includes a multidimensional array object, many derived objects (like masked arrays and matrices), and a variety of fast array operation routines, such as mathematical, logical,shape modulations, sorting, selecting, I/O, discrete Fourier transformation, basic linear algebra, basic statistical operation, random simulation and many more.

**Sklearn:**

Scikit-learn (sklearn) is a Python library that provides both unsupervised and supervised learning algorithms. It is the foundation for some of the technologies you might already be familiar with, such as NumPy, pandas, and Matplotlib.

**Python Imaging Library:**

Python Imaging Library is a free and open-source extension library for Python that adds support for opening, editing, and saving a wide range of image file formats. It works on Windows, Mac.

### 2.2 Modules:

**Tensorflow:**

TensorFlow is a free and open-source software library for data flow and differentiable programming that can be used to solve a variety of problems. It's a symbolic math library that's also used in neural networks and other machine learning applications. At Google, it is used for both research and development. The Google Brain team created the TensorFlow for internal Google use.

**Pandas:**

Pandas is an open-source Python library that use powerful data structures to provide high-performance data manipulations and analysis tools. Python was mostly used for data processing and munging. It made only a minor contribution to data analysis. This issue was solved by pandas. Regardless of the source of data, we can use Pandas to complete five common steps in data processing and analysis: plan, manipulate, model, and analyse. Python with Pandas is used in a variety of academic and commercial areas, including finance, economics, statistics, analytics, and more.

### 2.3 Features:

**Fastai:**

Fastai is a deep learning library with high-level components that can produce state- of-the-art results in traditional deep learning domains quickly and easily. It also offers low-level components that can be mixed and matched to create new methods for researchers.

**IPSDK:**

IPSDK is a C++ and Python image processing library. The library includes a full collection of image processing features for processing datasets, as well as a detailed and optimised set of 2D and 3D image processing functionalities. IPSDK adapts to the processor's architecture and capabilities automatically. Full PC cluster support, high performance and high availability computing, and other features are included in this library.

**Keras:**

Keras is high-level neural networks library written in Python that can run on top of either TensorFlow or Theano. The library was created with the goal of allowing fast experimentation. This deep learning library comes with a number of features, including support for both convolutional and recurrent networks, as well as simple and fast prototyping.

**Scikit-Image:**

Scikit-image is collection of algorithms for image processing.It includes algorithm for segmentation, geometric transformation, color space manipulation, analysis, filtering, morphology, feature detection. It is designed to inter operate with Python numericals and scientific libraries NumPy and SciPy.


Copyright @ 2021 Authors

**3.Algorithms:**

There are specific machine learning algorithms that are developed to handle complex real-world data problems. So, now we have seen the types of machine learning algorithms, let us study the top machine learning algorithms that exist and are actually used by data scientists.

**3.1 Naïve BayesClassifier Algorithm:**

What would have happened if you had to classify data texts such as a web page, a document or an email manually? Well, you would go mad! But thankfully this task is performed by Naïve Bayes Classifier Algorithm. This algorithm is based on Bayes Theorem of Probability and it allocates the element value to a population from one of the categories that is available.

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Where, y is class variable and X is dependent feature vector (of size n) where:

$$X = (x_1, x_2, x_3, ....., x_n)$$

An example of Naïve Bayes Classifier Algorithm usage is for Email Spam Filtering. Gmail use this algorithm to classify an email as Spam or Not Spam.

**3.2 K Means Clustering Algorithm:**

Let us imagine that you want to search the term "date" on Wikipedia. So Wikipedia groups the web pages that talk about the same ideas using K Means Clustering Algorithm (since it is popular algorithm for cluster analysis).

K Means Clustering Algorithm generally uses K number of clusters to operate on a given data set. In this manner, the output which contains K clusters with the input data partitioned among the clusters.

**3.3 Support Vector Machine Algorithm:**

Support Vector Machine Algorithm is used for classification or regression problems. In this, the data is divided into different classes by finding a particular line (hyper plane) which separates the data set into multiple classes. The Support Vector Machine Algorithm tries to find the hyper plane that maximizes the distance between the classes (known as margin maximization) as this increases the probability of classifying the data more accurately.

An example of Support Vector Machine Algorithm usage is for comparison of stock performance for stocks in the same sector. This help in managing investment making decisions by financial institutions.

**3.4 Apriori Algorithm:**

The Apriori Algorithm generate association rules using IF_THEN format. This means IF event A occurs, then event B also occurs with a certain probability.

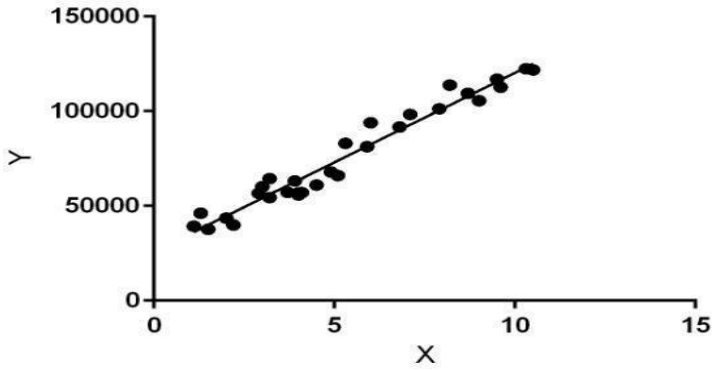
For example: IF a person buys car, THEN they will also buy car insurance. The Apriori Algorithm generates association rule by observing number of people who bought car insurance after buying a car.

An example of Apriori Algorithm usage is for Google auto-complete. When a word is typed in Google the Apriori Algorithm look for the associated words that are usually typed after the word and displays possibilities.

**3.5 Linear Regression Algorithm:**

The Linear Regression Algorithm shows the relationship between independent and a dependent variable. It demonstrates the impact on depend variable when independent variable is changed in any way.


Copyright @ 2021 Authors




Example of Linear Regression Algorithm is usage of risk assessment in insurance domain. Linear Regression analysis can be used for finding number of claims by customers of multiple ages and then deduce the increased risk as to the age of customer increases.

**3.6 Logistic Regression:**

The Logistic Regression Algorithm deals with discrete values where as Linear Regression Algorithm handles prediction in continuous values. So, Logistic Regression is suitable for binary classification where if an event occurs, it is classified as 1 and if not occurred it is classified as 0. Hence, the probability of a particular event occurrence is predicted based on given predictor variables.

**4. Output**:

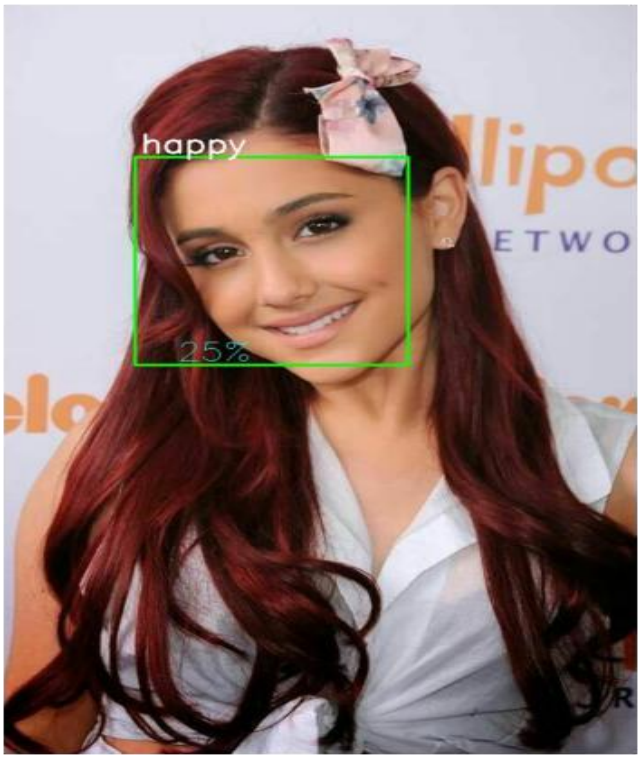



Figure 4(Result)

**5. Conclusion**

Machine learning is a promising field and with new research publishing every day. Today's AI require a lot of resources to train and produce accurate result. However there will be more human friendly AI systems in near future.


Copyright @ 2021 Authors

We can use this whole concept in knowing about patient's well being when he/she has been infected with a contagious disease. Which can be very useful in the present scenario?

**5.1 Why over project is better than the existing:**

As we take a look at all the different machine learning algorithms models by training the Datasets which are taken as input and considered 75% of data for training and 25% of data for testing and found accuracy of all models. And then we tried to boost the accuracy of models by using various models and also cleaning the datasets and we found out the best model which got the accuracy of 55.60% Recurrent Neural Networks for, 80% for Long Short term memory,. Finally, we used that model for finding the count of files of each of the emotions which we want to observe in these three datasets whereas the existing projects contained the accuracy of 80% and used to find the accuracy of each emotion either by taking datasets. So we can say that our project is better when compared to existing projects on the similar basis of computer vision out there.

**5.2 Drawbacks:**

The main drawback seems to be identifying faces when there are too many people present and the accuracy might go down a little bit.

We also need to figure out a way to constantly have a look at the Images to notice the changes in a patient.

**REFERENCES**

[1] https://en.wikipedia.org/wiki/Long_short-term_memory https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[2] https://machinelearningmastery.com/gentle-introduction-long-short-term-memorynetworks-experts/
[3] https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-stepexplanation-44e9eb85bf21
[4] https://www.researchgate.net/publication/325534924_A_Review_Paper_on_Human_Computer_Interaction
[5]https://ieeexplore.ieee.org/document/5593511
[6]https://www.engpaper.com/human-computer-interaction-2019.htm
[7]https://www.researchgate.net/publication/228475561_Understanding_and_predicting_human_behaviour
[8] https://www.researchgate.net/publication/308084190_A_deep_learning_approach_for_human_behavior_prediction_with_explanations_in_health_social_networks_social_restricted_Boltzmann_machine_SRBM
[9]https://fileserver.tk.informatik.tu-darmstadt.de/Publications/2007/Prediction.pdf
[10]https://www.kaggle.com/c/lish-moa
[11]https://www.csail.mit.edu/node/6315
[12]https://www.guide2research.com/journals/machine-learning
[13]https://www.nvidia.com/en-us/research/machine-learningartificial-intelligence/


Copyright @ 2021 Authors