ACKNOWLWDGEMENT

I would like to express my special thanks of gratitude to my Project Teacher "-----" for his valuable guidance and support in completing my Project.

I would also like to extend my gratitude to the MCA Coordinator "-----" to providing me with all the facilities that was required.

DATE:

<u>DECLARATION</u>

I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signature

Name of candidate

(-----)

MACHINE LEARNING WITH SQUID GAME DATASET

Abstract :

Machine learning is the scientific study of algorithms and statistic model where the machines are used to perform specific task without being explicitly programmed. It can predict any object or type which is manually done by human. In order to predict something, the computer need to be trained with training data. Here we have taken Squid Game dataset going to predict the Winner of the game by training the input dataset. The prediction can be done here with classification algorithms such as Logistic Regression. Then we after building the machine learning model it can able to predict the Winner correctly.

Keywords :

Logistic Regression , Squid Game Dataset , Machine Learning.

I.INTRODUCTION :

Squid Game dataset is the dataset we tried to made for prediction of the winner based on the games we provided them to play. The Participants have to play with there skills, strength, intelligence, etc. The one who lost any game in between so the player have to leave the game at that point. The dataset contains 63 participants overall. With this dataset as input our model is going to be developed with various classification algorithms.

And this model development has five stages to be processed before its prediction they are:

i Preparation of data.

ii Visualization of dataset.

iii Training and testing of dataset.

iv Choosing the classification algorithm to build the model.

v Evaluate the Model.

II.PROBLEM STATEMENT :

This data set consists of different parameters some of them are included for the model and some are not. For example the Name column in dataset is not considered because name cant help us to predict if the given person is winner pr not. Some of the parameters which are considered are age, strength, intelligence. The games players played were kabaddi, Hopscotch, Hide and Seek, Carrom, Pitthoo

we will be predicting the names of the winners. The data consists of continuous numeric values which describe the scores of the respective players. We will be training the model based on these scores.

The data is spitted into two parts:

1. Training data.
2. Testing data.

Once we have understood what the dataset is about, we can start training a model now we split our data into train and test data this then will be applied to various machine learning algorithms to predict the best results.

And the training data as input to the model and Testing data as output is Evaluated with its result.

III.PRE REQUISTES :

To implement our machine learning model, we need the following as the pre requisites to run our model successfully they are:

1. CPU core: i3 (minimum)
2. RAM: 4GB (minimum)
3. Jupyter Notebook
4. Necessary Packages

A. PREPARATION OF DATA :

The data is taken from Goggle Form which we made and distributed among our friends and it is imported to the Jupyter Notebook environment as a data frame using Python packages such as NumPy and Pandas. Using Pandas, the dataset can be viewed as a tabular format.

From the above Squid Game dataset, we can able to see the dataset with 3 features (Age, Strength and Intelligence) and 5 Different Games (Kabaddi, Hide and Seek, Hopscotch, Carrom, Pitthoo) which are the independent variable and the Survived is considered as target variable (dependent variable) where the actual prediction is going to be happening.
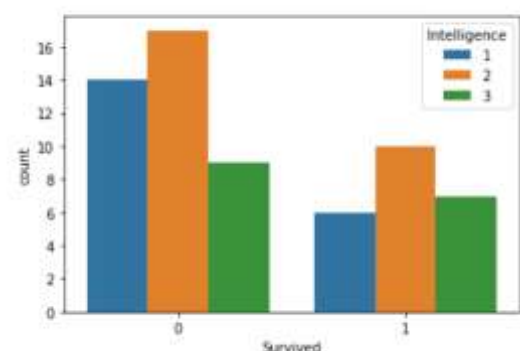
B. VISUALIZATION OF DATASET :

The data visualization is the process of visualizing the dataset in graphical form. Here we have visualized the dataset using the package called as heatmap, countplot and seaborn. The visualization is done for the dataset with different classes such as:
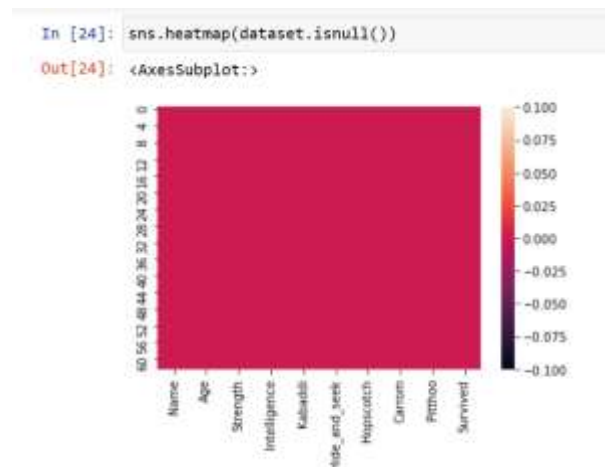
1. Intelligence :

```
intt = dataset['Intelligence']

sns.countplot(y,hue = intt)
```

2. Heatmap to see if there is any null values or not :

```
In [24]: sns.heatmap(dataset.isnull())
Out[24]: <AxesSubplot:>
```


C. TRAINING AND TESTING OF DATASET :

The dataset is spitted with training and testing dataset. Here the independent variable is separated from the dependent variable so that it can be used for test dataset to predict the result.

Training And Testing :

D. CHOOSING THE CLASSIFICATION :

Algorithm, we build the model using the classification algorithms from the Sci-kit learn package. Choosing the algorithm for developing the model is essential to get the prediction. The classification part using algorithm is all done after the data pre-processing, visualization, training and testing of dataset.

a. Logistic Regression Classifier.

Applying Logistic Regression:

So, as a start we have picked Logistic Regression Classifier as the algorithm to get the Winner prediction using target value which is trained with input. Logistic regression aims to solve classification problems. It does this by predicting categorical outcomes, unlike linear regression that predicts a continuous outcome. And it's found that Logistic Regression is predicting the Winner.

b. To Evaluate the Model :

Now we are in the final stage that is to evaluate the model by choosing any one or more classification algorithm which we have used. And the evaluation is done on target variable Species. Here we have used K-NN as the algorithm to evaluate our machine learning model.

```
In [47]: from sklearn.linear_model import LogisticRegression

In [48]: model = LogisticRegression()

In [49]: model.fit(X_train,y_train)

Out[49]: LogisticRegression()

In [50]: model.coef_

Out[50]: array([[ 0.04414644,  0.03317586,  0.09017272,  0.12809617,  0.13609782,
          0.        ,  0.19314524,  0.        ,  0.06530368, -0.27995643,
          0.15413661, -0.19003141,  0.        ,  0.22399445,  0.1154448 ,
         -0.62774031,  0.        , -0.01302069, -0.07737412,  0.25287071,
         -0.62699081, -0.00918351, -0.03052409, -0.40253797, -0.05959631,
          0.        , -0.0079208 , -0.07757823, -0.02586687, -0.1253988 ,
          0.08644141,  0.08078687,  0.        , -0.21685386, -0.50003184,
          0.15701433,  0.        , -0.14815727,  0.47067706,  0.        ,
         -0.07641068, -0.11304067, -0.51313951,  0.09096209,  0.33303185,
          0.15046747,  0.33033566,  0.56620098,  1.21940551,  1.5263478 ,
          5.33399859,  1.4498188 ]])

In [51]: y_pred = model.predict(X_test)
```

```
In [51]: y_pred = model.predict(X_test)

In [52]: y_pred

Out[52]: array([0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0],
         dtype=int64)

In [53]: import numpy as np

In [54]: y_pred_np = np.array(y_pred)

In [55]: y_pred_np

Out[55]: array([0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0],
         dtype=int64)
```

Here we have used Logistic Regression classification algorithm on Squid Game dataset to predict the result.

E.COMPARING THE ACCURACY FOR CLASSIFIERS :

After all the process our model is ready to predict the Winner.

Here We'll check for the error:

```
In [107]: mse = sklearn.metrics.mean_squared_error(y_test,y_pred)

In [108]: print(mse)
          0.0
```

The error is 0.0 which is perfect. Now we can conclude that our model provides with 100% accuracy.

F. DOMAIN :

The purpose of this project was to gain introductory exposure to Machine Learning Classification concepts along with data visualization.

The project makes heavy use of Scikit-Learn, Pandas and Data Visualization Libraries.

G.CONCLUSION :

In this paper we have implemented a machine learning model which can predict the winners in more easy way.

H.REFERENCES :

[1] https://towardsdatascience.com/python-for-data-sciencefrom-scratch-part-iii-7755f6defcc3

[2] https://docs.google.com/forms/d/e/1FAIpQLSdfvDc955OZMQtEP1HDVrdUpK2IOTxLZFyTEpl5ixt0_OpsRw/viewform?usp=sf_link