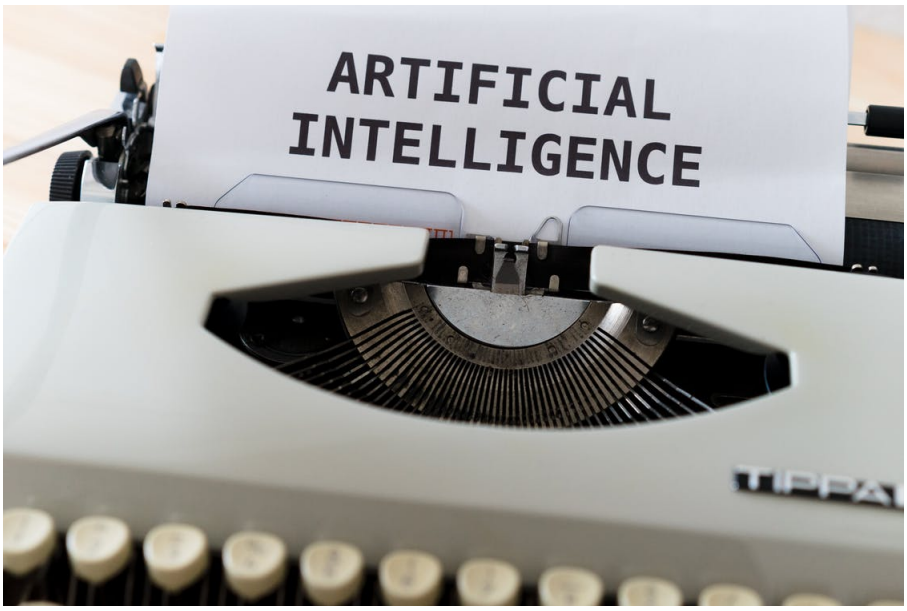
Group Assignment 1: Challenge - AI Environment Creation and Testing



Department: Computer Science

Facilitator: Kudakwashe Dandajena

Group Members: Diana Imali Luvandale, Heather Mataruse, Ian Wanjohi, Lavina Kathambi

1 Introduction

We used a dataset on malaria cases in African countries from the year 2007 - 2017. Below is a breakdown of how we used different libararies to manipulate the dataset and make observations.

2 Data-set

The data-set we used in our assignment includes data on the incidence of malaria in all African countries from 2007 to 2017. The incidence of malaria is a variable that counts the number of malaria cases per 1000 people in areas where malaria transmission occurs.

The malaria cases reported include the number of malaria cases that have been confirmed by examination. The data set also includes preventative measures that have been taken to prevent malaria, such as the use of insecticide-treated bed nets. The data was retrieved from the World Bank Open Data Source.

3 Libraries and Tools used

Find below the libraries and tools we used for this assignment as well as their descriptions:

- Numpy - NumPy, which is Numerical Python, is a library of multidimensional array objects and a set of array processing routines, and it also has functions for linear algebra, Fourier transforms, and matrix domain action.
- Pandas - Pandas is an open-source Python package that offers various data analysis tools and can present data in a way that is appropriate for data analysis through its sequence and data frame data structures.
- Scipy - One of the main packages forming the SciPy stack is the SciPy library. It offers many user-friendly and useful numerical routines, such as numerical integration and optimization routines, and it is also a Python library designed to work with NumPy arrays.

- Scikit-Learn - Scikit-learn is a Python library that is used to implement machine learning algorithms and has different algorithms such as vector machine support, random forests, and k-neighbors, and it also supports computational and scientific libraries such as NumPy and SciPy Python.
- Matplotlib - Matplotlib is a Python library used for plotting This python library provides objected-oriented APIs for incorporating plots into programs, which can be used in python scripts, shells, web application servers, and other graphical user interface toolkits.
- Keras - Keras allows users to increase the image while training the model, and it is an open-source software library that provides artificial neural networks with a Python interface and functions as a TensorFlow library interface.
- Tensorflow - TensorFlow allows developers to build graphs and frameworks for the data flow that explain how data moves through a graph or a series of nodes for processing. Each node in the graph represents a mathematical operation, and a multidimensional data array, or tensor, is each link or edge between nodes.
- Jupyter Notebook

4 Key Findings and Analysis

We used a dataset that shows data on Malaria cases in African countries from 2007-2017. We used Scipy Library to find the year with the most data on malaria cases which was 2007. We visualized our findings using the Matplotlib library. We used Numpy to try predicting number of malaria cases. We used Scikit-learn to split the data for testing and training of the model. We used Tensorflow to do basis manipulation of the data like adding and subtracting. We reaserched on Keras and found it was an open-source software library that provides artificial neural networks with a Python interface.

5 Group Members & Emails

1. Diana Imali Luvandale: d.luvandale@alustudent.com
2. Heather Mataruse: h.mataruse@alustudent.com
3. Ian Wanjohi: i.wanjohi@alustudent.com
4. Lavina Kathambi: l.murithi@alustudent.com

6 Conclusion

In conclusion, we were able to implement the above libraries on our data-set and were able to explore the various applications of scientific and machine learning libraries in our data.

7 References

1. The What's What of Keras and TensorFlow — upGrad blog (2019). Available at: https://www.upgrad.com/blog/the-whats-what-of-keras-and-tensorflow/ (Accessed: 24 January 2021).
2. Yegulalp, S. (2021) What is TensorFlow? The machine learning library explained, InfoWorld. Available at: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html (Accessed: 24 January 2021).
3. Kanani, B. (2019) Keras ImageDataGenerator with flow() - Machine Learning Tutorials, Machine Learning Tutorials. Available at: https://studymachinelearning.com/imagedatagenerator-with-flow/ (Accessed: 24 January 2021).
4. Python SciPy library (2021). Available at: https://www.python-ds.com/python-scipy (Accessed: 24 January 2021).
5. Johari, A. (2017) Matplotlib Tutorial — Python Matplotlib Library with Examples — Edureka, Edureka. Available at: https://www.edureka.co/blog/python-matplotlib-tutorial/ (Accessed: 24 January 2021).

6. S. and Jain, K. (2015) SKLearn — Scikit-Learn In Python — SciKit Learn Tutorial, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2015/01/learn-python-machine-learning-tool/ (Accessed: 24 January 2021).

7. Beginner's Tutorial on the Pandas Python Library (2021). Available at: https://stackabuse.com/beginners-tutorial-on-the-pandas-python-library/ (Accessed: 24 January 2021).

8. Scikit-learn Tutorial: Machine Learning in Python – Dataquest (2018). Available at: https://www.dataquest.io/blog/sci-kit-learn-tutorial/: :text=Scikit