The fastest way to build an accurate machine learning model.

“

Artificial Intelligence is a subset of Machine Learning.

“

Machine learning is a subset of Artificial intelligence.

“

Artificial intelligence and machine learning are the same.

“

Choose the odd one from the list below.

1. Andrew Ng
2. Yann LeCun
3. Kevin Hart
4. Geoffrey Hinton

“

A young data scientist named Ishmael discovered two binary classification models A and B that were accurate to 87.2% and 94.3%, respectively. He must make a decision. If you could guide him in making a decision, he would be very grateful.

“

```
In [6]: df['diagnosis'].value_counts()

Out[6]: B    357
        M    212
        Name: diagnosis, dtype: int64
```

```
In [14]: df['diagnosis'].value_counts()

Out[14]: B    212
         M    212
         Name: diagnosis, dtype: int64
```

A (94.3%) or B (87.2%) ?

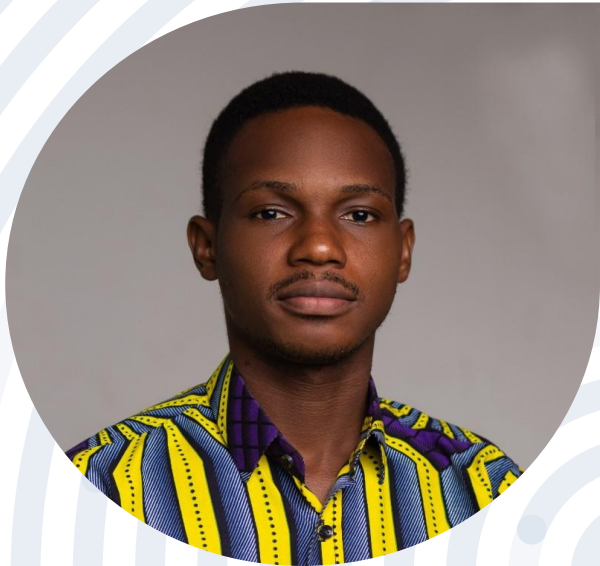
Introduction

I am Redeemer Salami Okekale.

Background: Biomedical Science, Data Science

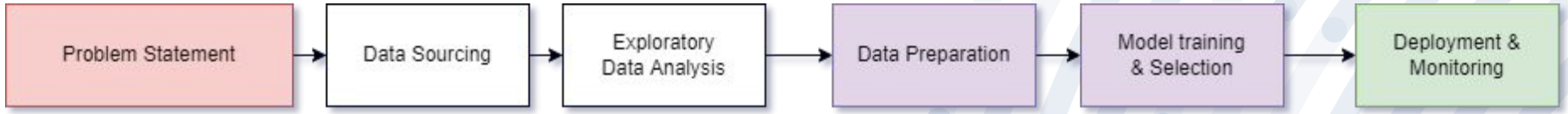
Industry: Healthcare, HealthTech, AI,

Interests: Medical research, Bioinformatics.

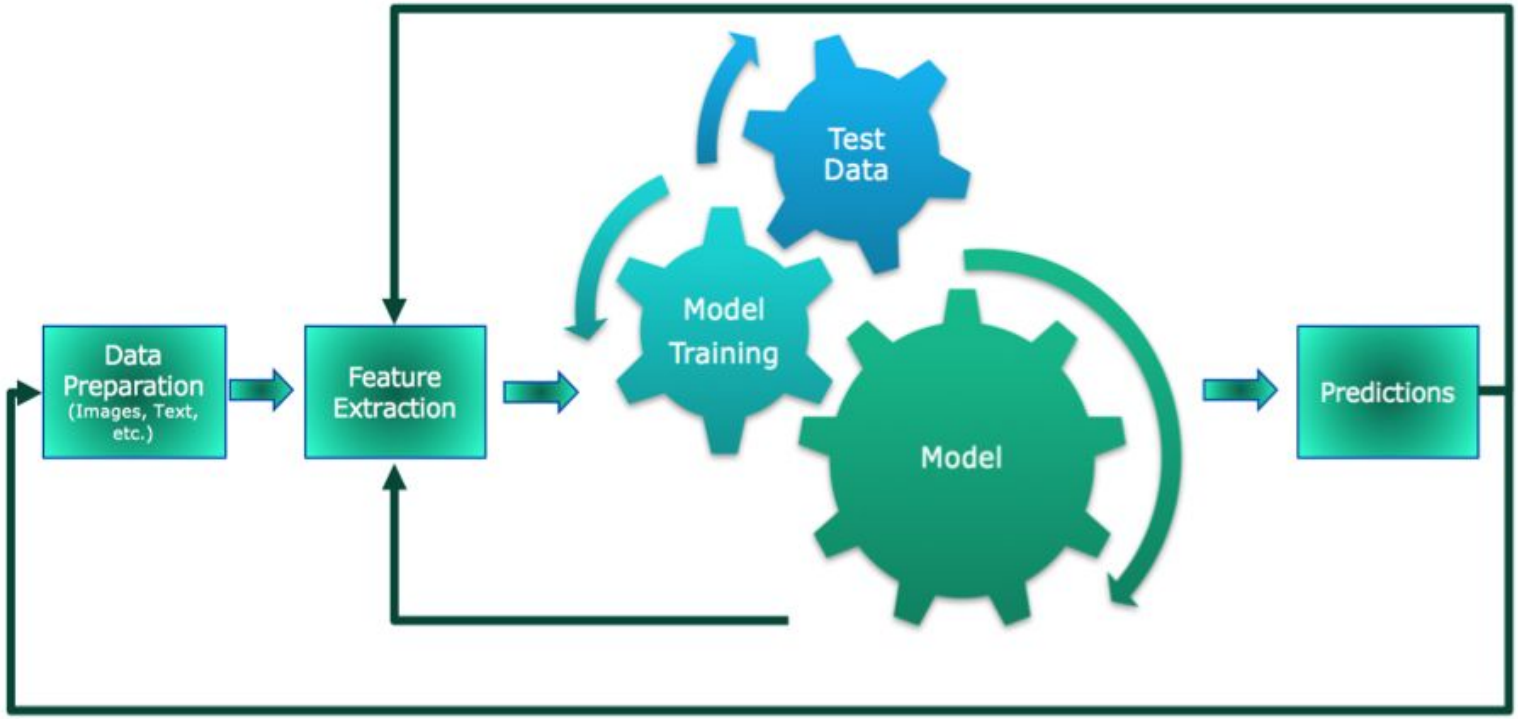


1. Machine Learning Life Cycle

ML Life Cycle



Data preparation, model training & selection



Challenges of the ML Life Cycle

- Time consuming
- Scalability Issues
- Cost
- Expertise

Pycaret

Moez Ali

ONCE YOU IDENTIFY THE PROBLEM, YOU ARE 50% CLOSER TO THE SOLUTION.

Types of Machine Learning models.

- Class -isfication
- Regression
- Clustering
- Anomaly detection etc.

Classification and Regression models

```
from pycaret.classification import *
train_data = 'train_data_csv'
test = 'test_data_csv'
s = setup(train_data, target='your_prediction_column')
best = compare_models()
evaluate_model(best)
predict_model(best, test)
```

```
[ ] from pycaret.regression import *
    train_data = 'train_data_csv'
    test = 'test_data_csv'
    s = setup(train_data, target='your_prediction_column')
    best = compare_models()
    evaluate_model(best)
    predict_model(best, test)
```

Clustering models

```
[ ]  from pycaret.clustering import *
     train_data = 'train_data_csv'
     s = setup(train_data, normalize=True)
     kmeans = create_model('kmeans')
     #apply model to data
     kmeans_results = assign_model(kmeans)

     #test
     new_data = pd.read_csv('new_data')
     predict_model(kmeans, data = new_data)

     #save
     save_model(kmeans, 'kmeans_pipeline')
```

IT'S PRACTICALS TIME!

Thanks!

Any questions?

Email: sredeemer24@gmail.com

Phone: 0555225870

LinkedIn: https://www.linkedin.com/in/redeemer-salami-okekale-a69a0215b/

References

1. **Kaggle Dataset:** https://www.kaggle.com/datasets/vijayaadithyanvg/breast-cancer-prediction
2. **Pycaret:** https://pycaret.gitbook.io/docs/
3. **Github:** https://github.com/oMarquess/pycon_2022.git