Genetic Algorithm in Optimization

By

Namya Vyas

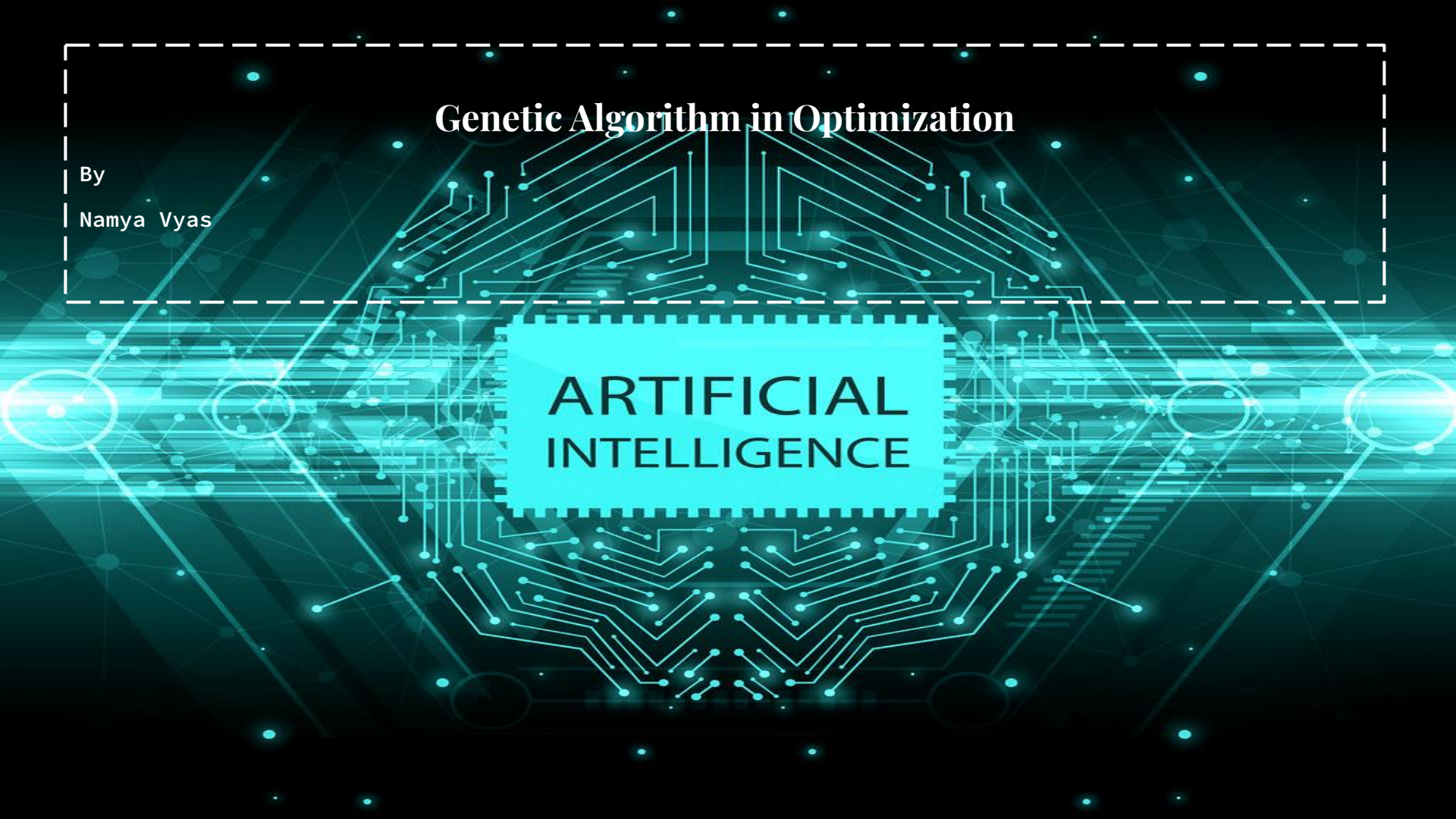


Table of contents:

Problem Description

Problem Definition

- Rainfall forecasting system based on genetic algorithm (GA).
- Aim of the project → Feature selection using Genetic algorithm in a binary classification model of rainfall prediction.
- Pre-optimization will be used, then the data will be fed to the prediction model.
- The model can be used anywhere such as, tourism planning, crop management, harvest seasons etc.

Motivation

- Advantage of the randomness in data; unlike other predictive algorithms
- new method that may not be discovered through typical statistical or regression methods.
- Not a simple random walk; exploit historical information to speculate on a new search point
- Expected improvements to be seen

Genetic Algorithm Introduction

Survival of Fittest

Genetic algorithms are randomized search algorithms that generate high-quality optimization solutions by imitating the biologically inspired natural selection process such as selection, cross-over, and mutation.

Terminology and Operators :-

Population:

Set of possible solutions for the stochastic search process to begin.

Chromosome(Genotype):

represents one candidate solution present in the generation or population.

Phenotype:

decoded parameter list for the genotype that is processed by the Genetic Algorithm

Fitness Function:

evaluates the individual solution or phenotypes for every generation to identify the fittest members.

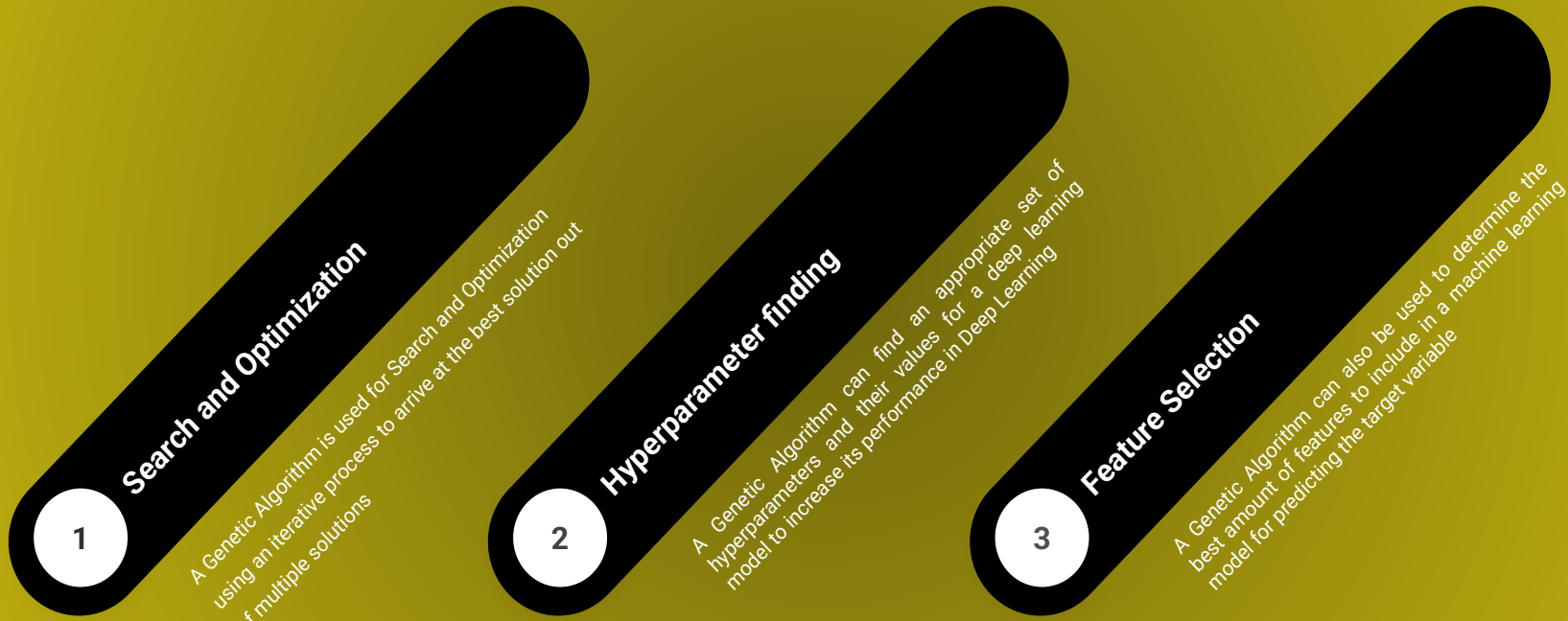
Selection:

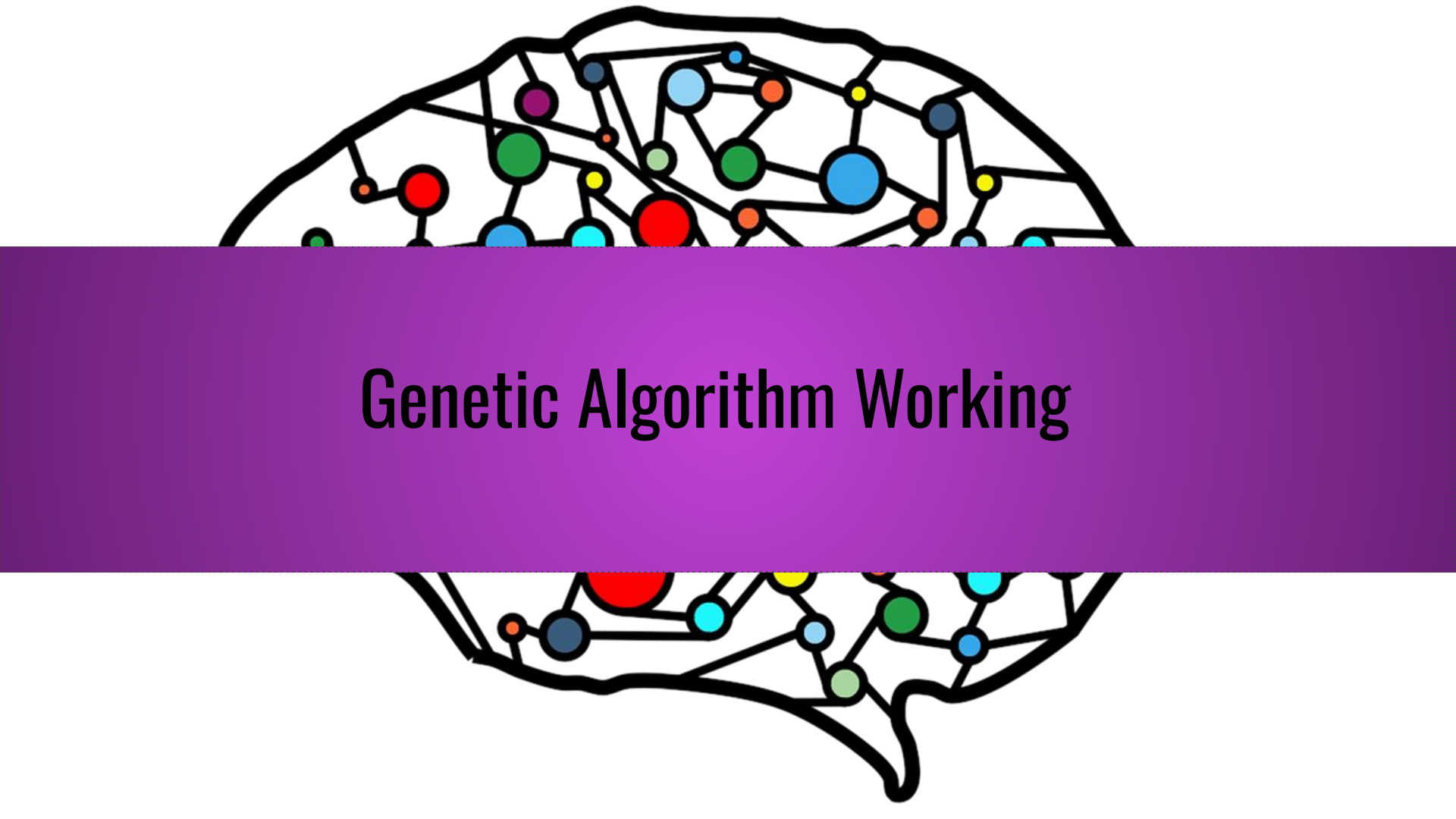
process of selecting the fittest solution from a population, and then the fittest solutions act as parents of the next generation of solutions.

Cross-over:

genes from the two fittest parents are randomly exchanged to form a new genotype or solution.

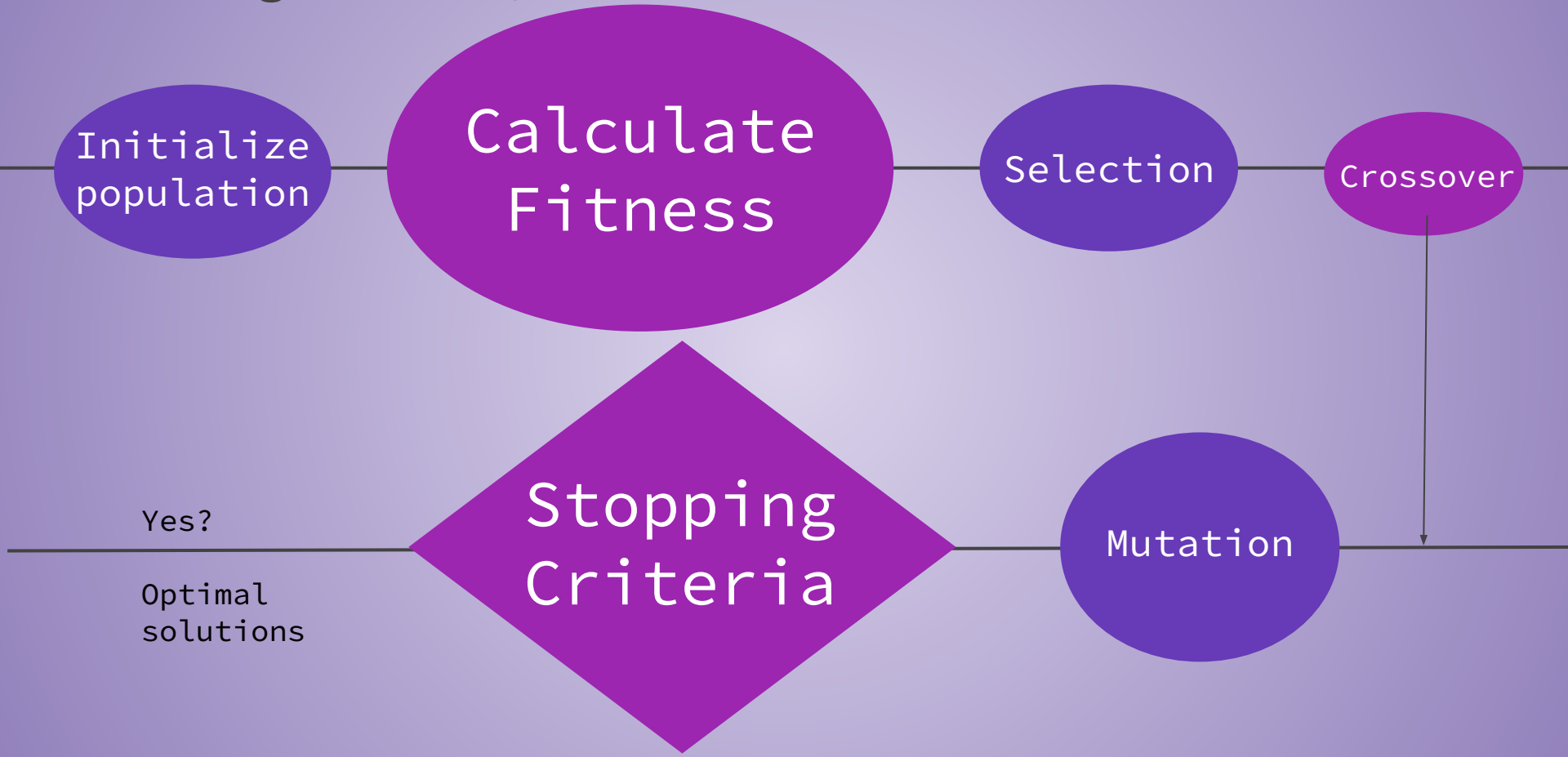
Applications of Genetic Algorithm

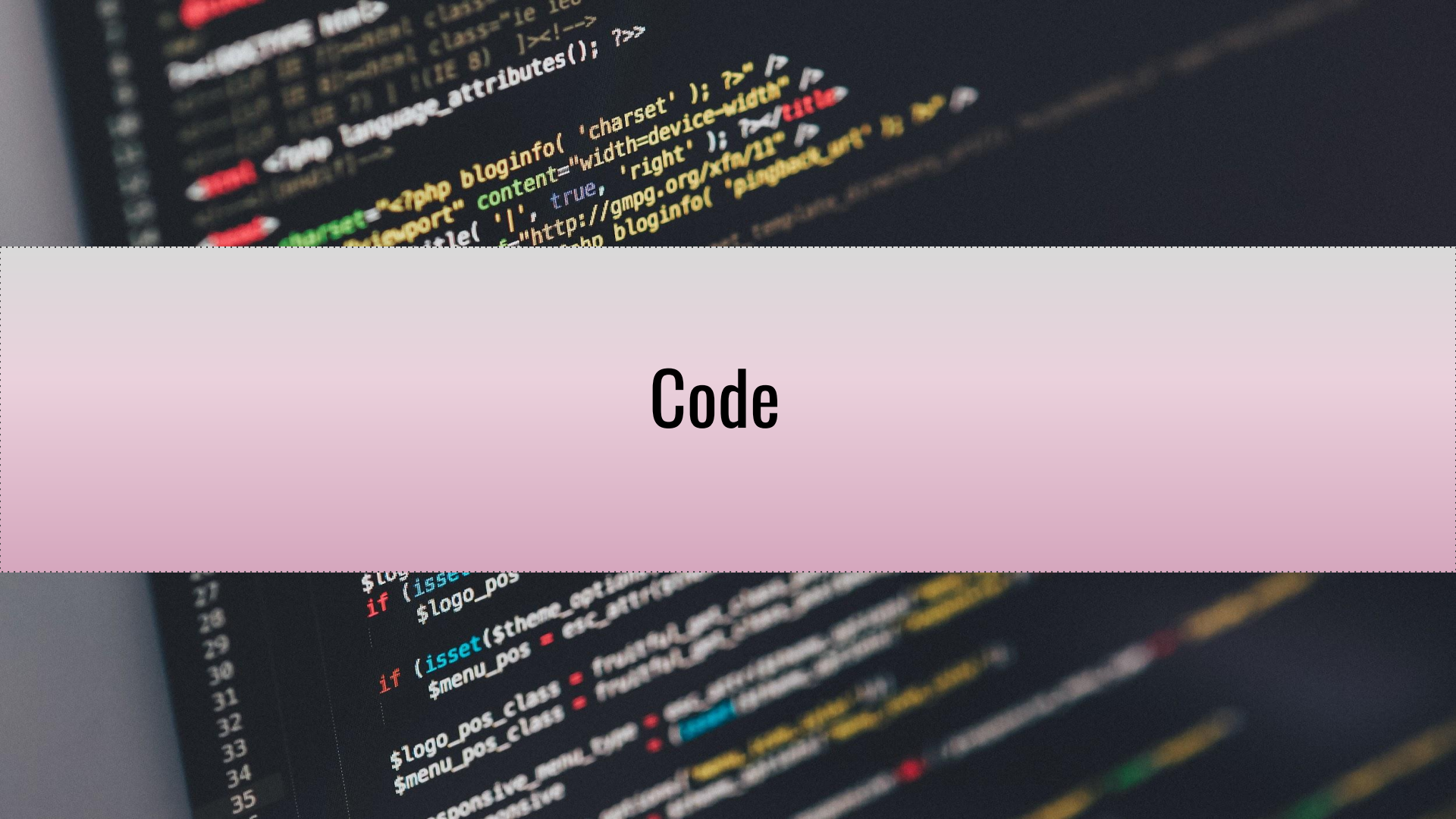




Genetic Algorithm Working

Genetic Algorithm Layout:




Code

About Dataset

Predict next-day rain by training classification models on the target variable RainTomorrow.

This dataset contains about 10 years of daily weather observations from various weather stations across Australia.

RainTomorrow is the target variable to predict. It means -- did it rain the next day, Yes or No? This column is Yes if the rain for that day was 1mm or more.

Source: Kaggle; http://www.bom.gov.au/climate/dwo/, http://www.bom.gov.au/climate/data.

Notebook

https://colab.research.google.com/drive/13qmGH8imEiTMzN3IWmQKO-WY_fS5cOeV?usp=sharing

Expected Outcome

Towards the aim —>

Expectation

- We use optimization for increasing the performance of the classifier and thereby making a rainfall prediction system.
- This set is such that it is the most optimized combination and extremely random

Result and Conclusion

Output

- ★ 1-2 % increase in the improvement can be seen with the generations when Knn and RadialSVM classifier is used for the prediction of rainfall.
- ★ The given dataset goes under overfitting with following classifiers:LinearSVM, Logistic, Random Forest, AdaBoost and Decision tree.

```
# import KNeighbors ClaSSifier from sklearn
from sklearn.neighbors import KNeighborsClassifier


# instantiate the model
knn = KNeighborsClassifier(n_neighbors=4)


# fit the model to the training set
logmodel= knn.fit(X_train, Y_train)
logmodel
```

```
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=4)
```

```
[ ] X_train,X_test, Y_train, Y_test = split(data,label)
    chromo_df_bc,score_bc=generations(data,label,size=80,n_feat=data.shape[1],n_parents=64,mutation_rate=0.20,n_gen=5,
                            X_train = X_train,X_test = X_test,Y_train = Y_train,Y_test = Y_test)
```

```
Best score in generation 1 : [0.8384]
Best score in generation 2 : [0.84208]
Best score in generation 3 : [0.84928]
Best score in generation 4 : [0.842]
Best score in generation 5 : [0.84008]
```

Improvement of 1-2% can be seen as the generation increases.

Presentation By:

Namya Vyas
20BIT101

Thank you!