



<hacker-ramp/>

we for the

Team Name :lvl_101

Team Details: Uggumudi Sai Lasya Reddy(Team Leader)

+917975448163

Leena Rai Kamarthy

+919113015524

Siriki Vagdevi

+918332053328

Data Classification

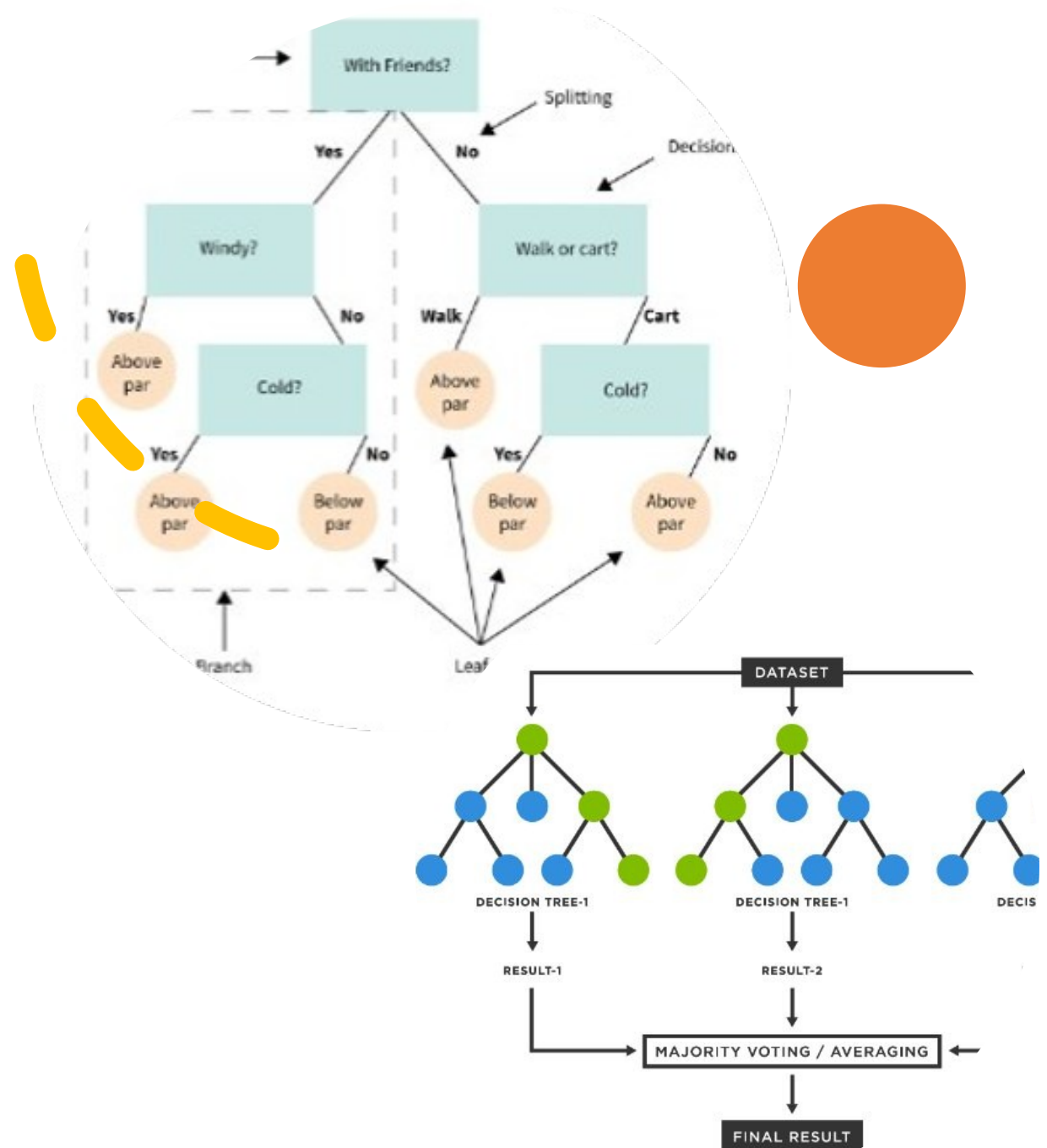
- The information is loaded by using the Pandas library.
- Dataset has **several columns** such as Gender, Master Category, Sub Category, Article Type, Base Colour, Season , Usage, Year and Product Display Name.
- Data Preprocessing : In this step **Label encoding** is applied to convert categorical variables into numerical form. By dropping rows with null values , missing values are handled.
- Feature Selection and Filtering : **Features are selected for model training**, and the **target variables** are **identified** after which data is filtered to retain only those samples that meet a minimum frequency threshold for certain categories . Target variables can be say articletype and usage .
- Data splitting : The data is divided into training and testing sets so as to evaluate the performance of the model .

Decision Tree Classifier :

- The Decision tree classifier is imported for developing a decision tree model. The function accuracy score is imported to evaluate the accuracy of the model predictions.
- An instance of the Decision tree classifier is created. It will be used to train this object articletype,usage .
- The **fit method** trains the decision tree model using the training dataset .
- The **predict method** creates forecasters for the test dataset and stores predicted labels.
- To calculate its accuracy, compare predicted labels with true labels using an accuracy function that passes its output

Random classifier :

- RandomForestClassifier is imported for creating a **random forest model** . **GridSearchCV** is imported for performing **hyperparameter tuning** through **cross-validation**. An instance of RandomForestClassifier is created, which will be used for training.






- Different values for **n_estimators** and **max_depth** are defined in a **parameter grid** for exploring purposes.
- The fit method does a grid search over defined hyperparameters to find the **best combination** of these that make up the model trained on training data. It extracts and stores the **best performing model** according to the grid search as **best_rf_articleType**.
- Using **predict() method** test dataset is predicted with best model. The accuracy of predictions is calculated as true labels compared with predicted labels and stored.

XGB Classifier:

- Implementation of **gradient boosted decision trees** designed for **speed and performance**. This line **trains the classifier using training data along with their respective labels**.
- For prediction accuracy, this line compares predicted labels against actual ones . The proportion of correctly predicted instances can be computed



The Project

Main Advantage	The model is flexible
Input V/s Prediction 	I/p-Age, Gender, Season O/p-Type of clothes / color or Input can include usage like casual or party and corresponding output can be given
Integration 	Because of its flexibility, the model can be integrated into Myntra's search or Myntra's filters or a specific section called trends can be added.
Improvement 	The model can still be improved a lot with a good dataset . <u>XGBoost</u> Engineering new features- like year + season or gender + age

General Benefits

- **Personalized Recommendations:** The model can forecast on fashion item attributes in view of the user's preference to enable it suggest similar items that might be of interest to the users.
- **Improved Search:** One can search for item characteristics and the model will ensure that search results are relevant and accurate.
- **Inventory Management:** Knowing the trend by predicting popular attributes would help manage inventory by adjusting stock levels.
- **Improved User Experience:** It can make better recommendations for user preferences, thus generating a captivating and individualized shopping experience.
- **Increased Sales:** Since recommendations are personalized and search findings optimized, users are likely to be satisfied, leading to an increase in buying propensity.
- **Customer Retention:** The improved customer involvement through enhanced user experience is expected to enhance customer loyalty and prevent loss of customers through personalization.
- **Competitive Edge:** Being able to offer a much more sophisticated shopping platform as well as being one step ahead after embracing latest techniques in machine learning.