

# Lab Exercise 4 – Supervised Learning

## Naïve Bayesian, Decision Tree, and SVM

### Step 1:

Go to <https://jupyter.org/> and install Jupyter Lab.

The screenshot shows the official Jupyter website at jupyter.org. The page features a dark header with the URL and a navigation bar with links like Bookmarks, MFU, DE, BDA, Golden Triangle, GMAIL, Outlook, Maps, YouTube, Login, Amazon, ChatGPT, and RelaX. Below the header is a banner with the text "Free software, open standards, and web services for interactive computing across all programming languages". A large central image displays the JupyterLab interface, which includes a file browser on the left, a code editor in the center, and various data visualizations and notebooks on the right. To the right of the interface, there is descriptive text about JupyterLab and two buttons: "Try it in your browser" and "Install JupyterLab".

### Step 2:

Create a folder "Jupyter Examples" in C:\ drive and execute **Jupyter Lab**  
C:\Users\Admin\Jupyter Examples>jupyter lab

```
C:\Windows\System32\cmd.exe - jupyter lab
Microsoft Windows [Version 10.0.19045.5371]
(c) Microsoft Corporation. All rights reserved.

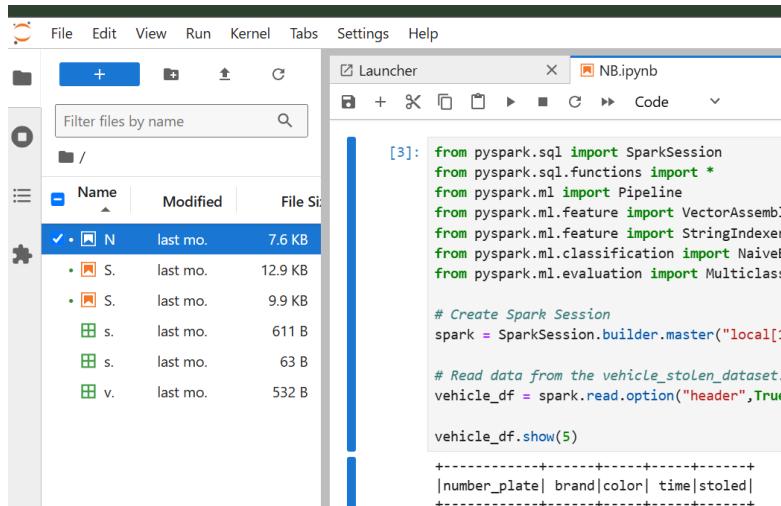
C:\Users\Admin\Jupyter Examples>jupyter lab
[I 2025-01-21 11:11:09.752 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2025-01-21 11:11:09.763 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2025-01-21 11:11:09.777 ServerApp] jupyterlab | extension was successfully linked.
[I 2025-01-21 11:11:09.788 ServerApp] notebook | extension was successfully linked.
[I 2025-01-21 11:11:10.410 ServerApp] notebook_shim | extension was successfully linked.
[I 2025-01-21 11:11:10.458 ServerApp] notebook_shim | extension was successfully loaded.
[I 2025-01-21 11:11:10.461 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2025-01-21 11:11:10.463 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2025-01-21 11:11:10.467 LabApp] JupyterLab extension loaded from C:\Users\Admin\AppData\Local\lib\site-packages\jupyterlab
[2025-01-21 11:11:10.467 LabApp] JupyterLab extension loaded from C:\Users\Admin\AppData\Local\lib\site-packages\jupyterlab
```

You should see the following site:

The screenshot shows a web browser window at localhost:8889/fab/tree/NB.ipynb. The browser has a dark theme and includes a navigation bar with links to Bookmarks, MFU, DE, BDA, Golden Triangle, GMAIL, Outlook, Maps, YouTube, Login, Amazon, ChatGPT, and RelaX. Below the navigation bar is a toolbar with File, Edit, View, Run, Kernel, Tab, Settings, and Help. The main area of the browser shows a file browser on the left listing files like NB.ipynb, Spark-SQL.ipynb, and SVM.ipynb. On the right, there is a code editor window displaying Python code related to PySpark and machine learning, specifically Naive Bayes and MulticlassClassificationEvaluator. The code includes imports for SparkSession, SQL functions, ML Pipeline, VectorAssembler, StringIndexer, and Evaluation metrics, along with code to read a dataset and show its first few rows.

### Step 3:

Copy and paste the code from the file: NB.ipynb into a Jupyter notebook.



```
[3]: from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.ml import Pipeline
from pyspark.ml.feature import VectorAssembler
from pyspark.ml.feature import StringIndexer
from pyspark.ml.classification import NaiveBayes
from pyspark.ml.evaluation import MulticlassClassificationEvaluator

# Create Spark Session
spark = SparkSession.builder.master("local[1]").getOrCreate()

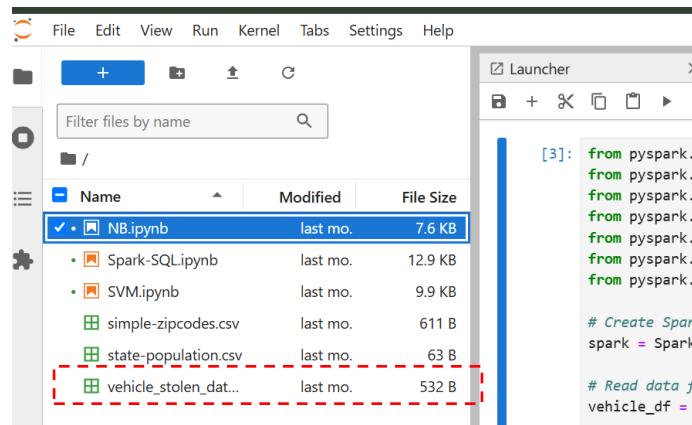
# Read data from the vehicle_stolen_dataset.
vehicle_df = spark.read.option("header",True).option("inferSchema",True).csv("vehicle_stolen_dataset.csv")

vehicle_df.show(5)

+-----+-----+-----+-----+
|number_plate| brand|color| time|stolen|
+-----+-----+-----+-----+
```

#### Step 4:

Make sure to keep the dataset file **vehicle\_stolen\_dataset.csv** in the working folder.



```
[3]: from pyspark.

# Create Spark Session
spark = SparkSession.builder.master("local[1]").getOrCreate()

# Read data from the vehicle_stolen_dataset.
vehicle_df = spark.read.option("header",True).option("inferSchema",True).csv("vehicle_stolen_dataset.csv")
```

#### Step 5:

Load the next two files *spark-SQL.ipynb* and *SVM.ipynb*. The dataset for *spark-SQL.ipynb* is *simple-zipcodes.csv*. For *SVM.ipynb*, there is no need to use any *csv* file, because *breast\_cancer()* dataset is downloaded from *sklearn.datasets*.

#### Step 6:

Record the accuracy for each program after executing the code.