National College of Ireland

Evaluating the Sensitivity of Machine Learning Algorithms to Training Data Size in OS X and Memory Malware Detection

MSc Research Project
MSc in Data Analytics

Student ID: x23189428

School of Computing

National College of Ireland

Supervisor: Vikas Tomer

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Devika Tamidala
Student ID:	X23189428
Programme:	MSc in Data Analytics **Year:** 2024
Module:	Research Project
Supervisor:	Vikas Tomer
Submission Due Date:	12/12/2024
Project Title:	Evaluating the Sensitivity of Machine Learning Algorithms to Training Data Size in OS X and Memory Malware Detection
Word Count:	547 **Page Count:** 4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Devika Tamidala

Date: 12/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Devika Tamidala

X23189428

Introduction

This configuration manual provides details about hardware, software, programming environment and libraries that are needed for implementing and testing machine learning algorithms for malware detection. Environmental Setup

Hardware requirements

- **Processor**: Intel Core i5 (or equivalent) and above.
- **RAM**: 8 GB or more if required

Software requirements

- Python programming of version 3.9 or above
- Jupyter notebook
- Google Colab (If required)
- Microsoft Excel for datasets
- Microsoft word for the documentation

Visualization Tools

- Python visualization using libraries like Matplotlib and Seaborn.

Machine Learning and Model Development

- **Scikit-learn** (Logistic Regression, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (GNB) models)
- Metrics for evaluation (accuracy, precision, recall, and F1-score).
- RandomizedSearchCV for hyperparameter optimization.
- **Imbalanced-learn**: Applied SMOTE (Synthetic Minority Oversampling Technique) for balancing the OS X dataset.

Data Manipulation and Analysis

- Pandas: Handling and processing datasets, Splitting datasets into training, validation, and testing sets.
- NumPy: Efficient numerical computations and matrix operations.

Tools Setup

- Anaconda Navigator: For the creation of Load Balancing and fault tolerance program.

Installing Anaconda Navigator

Step 1: Visit the Official Anaconda Website:

Open a web browser and navigate to the Anaconda Download Page.

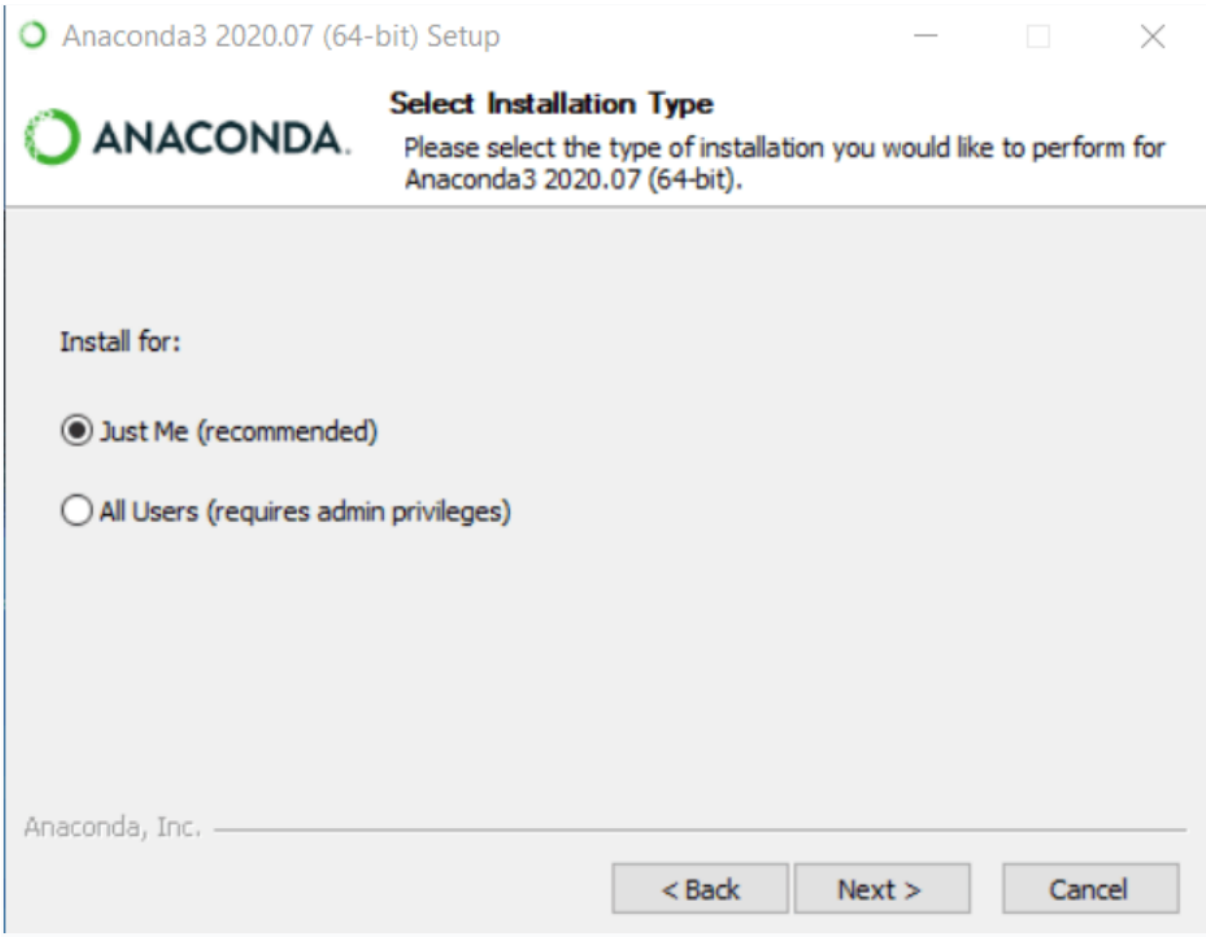
https://www.anaconda.com/products/navigator

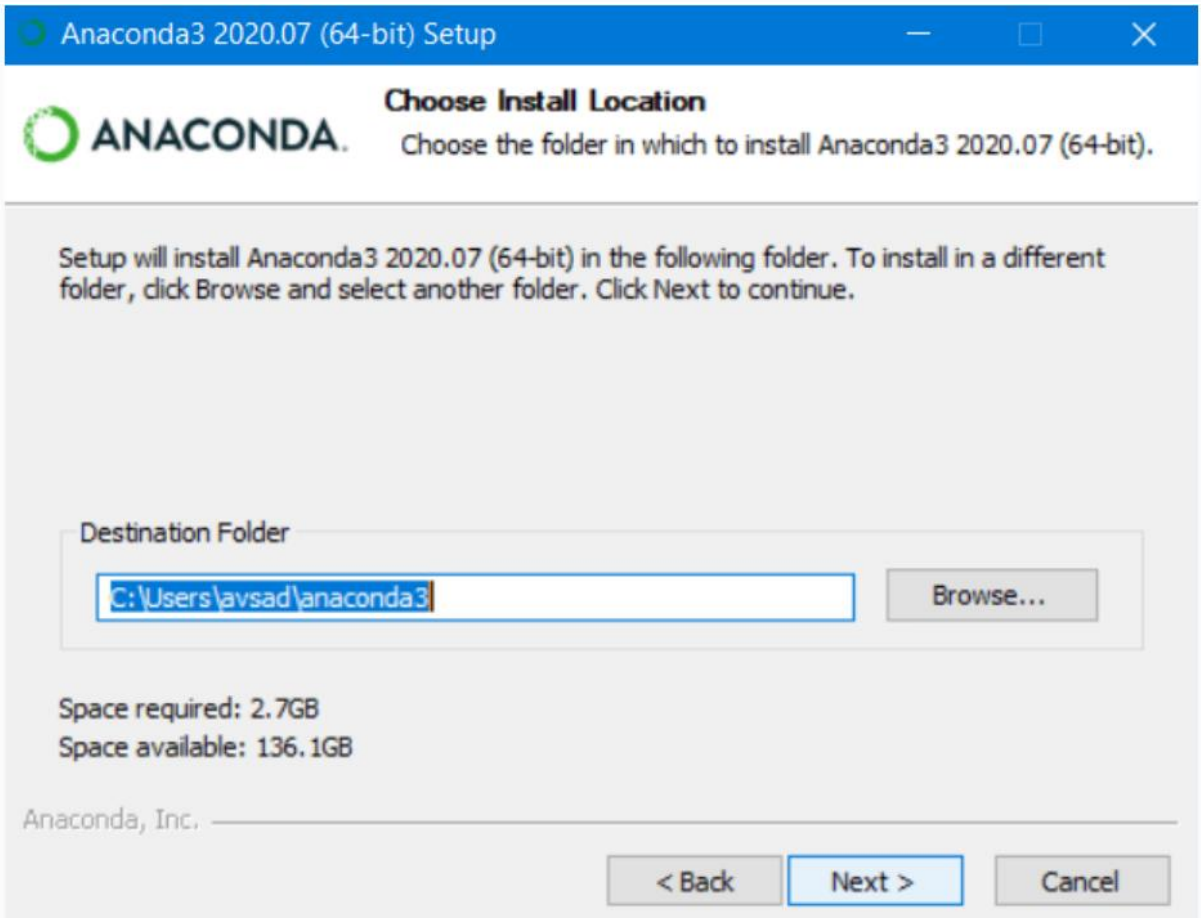
Step 2: Choose Your Operating System

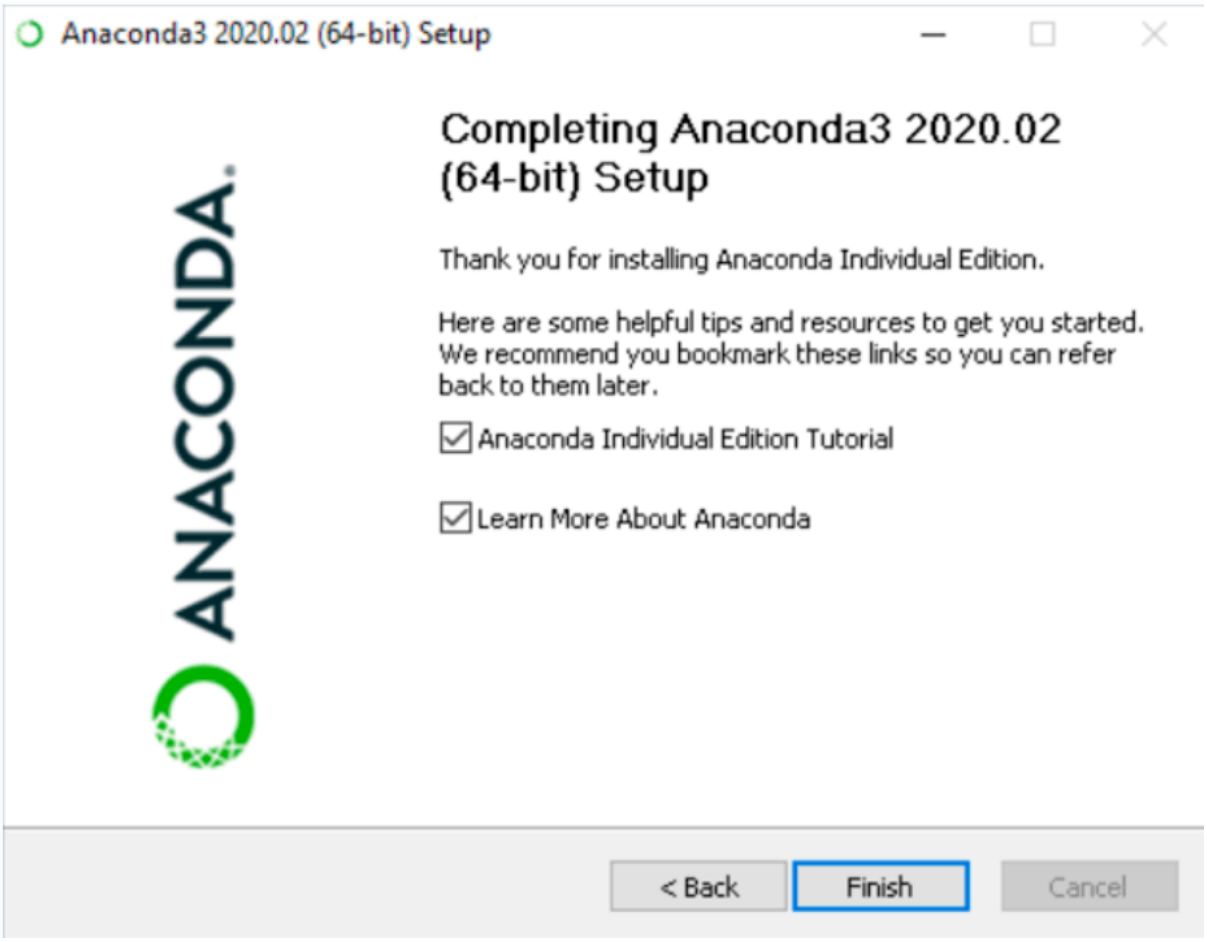
Select Windows operating system

Step 3: Download the Installer

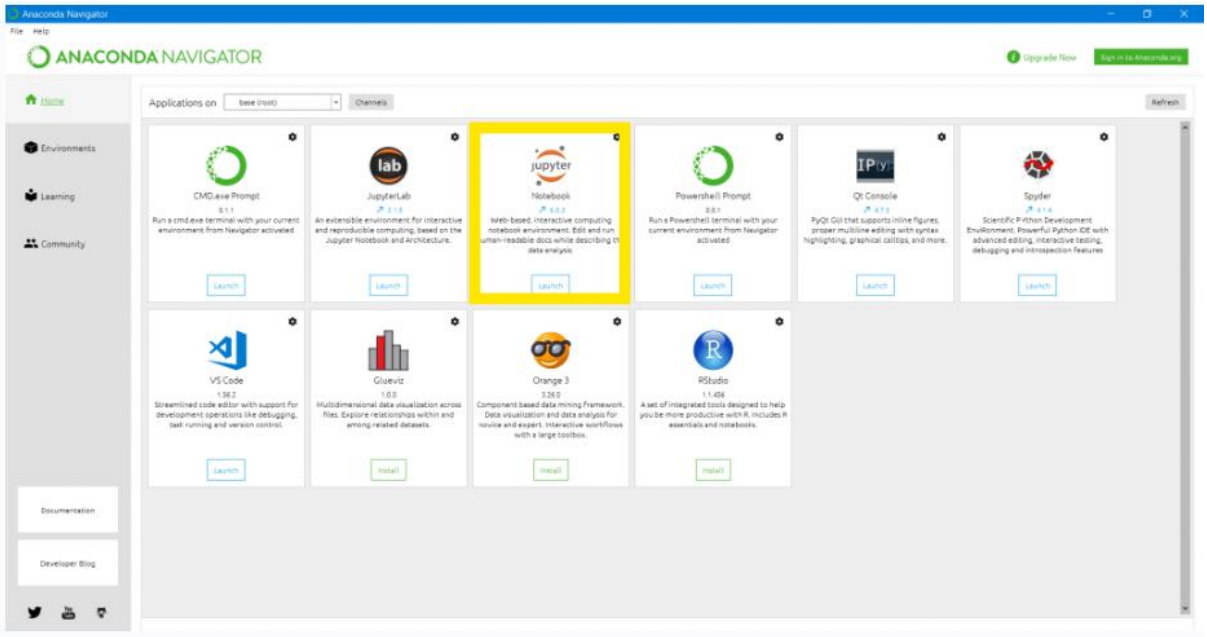
Once the operating system is selected, the page will provide the appropriate installer. Click the download link to begin downloading the installer file.


Anaconda3 2020.07 (64-bit) Setup
ANACONDA.
Select Installation Type
Please select the type of installation you would like to perform for Anaconda3 2020.07 (64-bit).
Install for:
Just Me (recommended)
All Users (requires admin privileges)
Anaconda, Inc.
< Back
Next >
Cancel


Anaconda3 2020.07 (64-bit) Setup
ANACONDA.
Choose Install Location
Choose the folder in which to install Anaconda3 2020.07 (64-bit).
Setup will install Anaconda3 2020.07 (64-bit) in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Destination Folder
C:\Users\avsad\anaconda3
Browse...
Space required: 2.7GB
Space available: 136.1GB
Anaconda, Inc.
< Back
Next >
Cancel



Step 4: Once installed, open Anaconda Navigator. Within Anaconda Navigator, launch Jupyter Notebook.



Step 5: You can use any browser of your choice to open and work with Jupyter Notebook.