|  |
| --- |
| Assessment Report  subtitle text here |
| Company Name Street Address  City, ST ZIP Code  Phone  Email |



## Relevant Software Tools Used For This Assessment

### JupyterNotebooks

Jupyter is a project and community whose purpose is to support interactive data science and scientific computing across all programming languages.

In this project it is used for the writing and plotting of graphs because it lets you see the results of your code on the go.

Below are the pros (appropriateness) of jupyter.

1. It’s easy to use other people’s work as a starting point. You can run cell by cell to better get an understanding of what the code does. Which in most cases would save the developer a lot of time and energy.
2. Very easy to host server side, which is useful for security purposes. A lot of data is sensitive and should be protected, and one of the steps toward that is no data is stored on local machines. A server-side Jupyter Notebook setup gives you that for free.
3. They're great for showcasing your work. You can see both the code and the results.

Below are the cons (limitations) of jupyter.

1. It is almost impossible to enable good code versioning.
2. The non-linear workflow of jupyter.
3. Jupyter is bad for running long asynchronous tasks.

#### Python

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python libraries like pandas, numpy, and matplotlib are used in this project.

Below are the pros (appropriateness) of python.

1. Vast libraries support. Which makes it easier to accomplish anything in it, Projects in categories like web development, data science, and artificial intelligence are easier and faster to develop due to already developed libraries in python for them.
2. Easy to read and write. Pythons design philosophy emphasizes code readability with its use of significant indentation. Unlike java, objective C, etc. python stands out as the best language when it comes to readability.
3. Dynamically typed? Yes, it is. Python is a dynamically typed language. What is dynamic? We don’t have to declare the type of variable while assigning a value to a variable in python. Other languages like C, C++, Java, etc…, there is a strict declaration of variables before assigning a value to them. Python don’t have any problem even if we don’t declare the type of variable. It states the type of variable in the runtime of the program. So, Python is a dynamically typed language.

Below are the cons (limitations) of python.

1. Runtime errors. A program with a runtime error is one that passed the interpreter’s syntax checks, and started to execute. However, during the execution of one of the statements in the program, an error occurred that caused the interpreter to stop executing the program and display an error message. Runtime errors are also called **exceptions** because they usually indicate that something exceptional (and bad) has happened. One of the major drawbacks of this language is that its design has numerous issues. Python programmers face several issues regarding the design of the language. This language requires more testing and also it has errors that only show up at runtime this is because the language is dynamically typed.
2. Slow speed. Although Python is easy to learn because of its syntax and being a dynamically typed language, it is simply too slow. The line by line execution of code often leads to **slow execution**.

The dynamic nature of Python is also responsible for the **slow speed**of Python because it has to do the extra work while executing code. So, Python is not used for purposes where speed is an important aspect of the project.

1. Not memory efficient. To provide simplicity to the developer, Python has to do a little tradeoff. The Python programming language uses a **large amount of memory**. This can be a disadvantage while building applications when we prefer memory optimization.

#### Pandas

#### Pandas is a software library written for the python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. In this project pandas is mainly used for data cleansing and processing.

Below are the pros (appropriateness) of pandas.

1. Data representation. This helps the developer in seeing a bit of the data he/she is working with, without having to look through the entire data. For instance, the ‘head’ method shows the developer top four of the entire data. Or the amount passed in as a parameter, this makes it easier to get a hint or little view of the data he/she will be working with
2. Makes data flexible and customizable.
3. Less writing and more work done.

Below are the cons (Limitations) of pandas.

1. Bad documentation. Its documentation is a lot complex to work through as a beginner in data science.
2. Steep learning curve. Pandas can be a lot complex for fresher in python development.
3. Poor compatibility for 3D matrices.

#### Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. Matplotlib is used in this project for graph plotting.

Below are the pros (appropriateness) of matplotlib.

1. Provides control to various elements of a figure such as DPI, figure colour, figure size.
2. It provides high-quality images and plots in various formats such as png, pdf, pgf, etc.
3. Its cross-platform and portable.

Word 2013

Word 2013 is **a word processing application that allows you to create a variety of documents like letters, flyers, and reports**. In this project it is used for all documenting, including this report.

Below are the pros (appropriateness) of Word 2013.

1. Availability. Word 2013 is available all over the world, and also compatible for almost all operating systems.
2. Document Flexibility.
3. Integration with office programs.

Below are the cons (limitations) of Word 2013.

1. Complexity.
2. Cost.

The Security Other Risk Implications Related To The Utilization Of Software Libraries And Tools To Real World Applications In The Context Of DS, AI, Or Fintech.

When it comes to building real world applications in the context of data science, artificial intelligence, or fintech in python, libraries like BeautifulSoup (Data mining), numpy, pandas, sciPy, scikit-learn (Data processing and modeling), and matplotlib (data visualization) are mostly used for its development.

Below are the outlined security and other risks related to it.

#### NumPy

#### Numpy is a library for the python programming used in scientific computing. The current version of numpy relies on insecure default usage of a python module that could lead to remote code execution. This bug affects numpy versions 1.10 and later. The popular library is working to deliver a fix to the bug.

#### According to <https://cyware.com/> a security researcher sherwel nan reported the bug [(cve-2019-6446)](https://nvd.nist.gov/vuln/detail/CVE-2019-6446)on January 16, 2019. Nan stated that if a python application loads malicious data via the 'numpy. Load' function, an attacker could obtain remote code execution on the machine.

The bug CVE-2019-6446 is marked as critical, with a severity score of 9.8 as per the Common Vulnerability Scoring System (CVSS) version 3.

**Pickle Module**

The problem is with the ['pickle'](https://pythontips.com/2013/08/02/what-is-pickle-in-python/) module, which is used for transforming Python object structures into a format that can be stored on disk or in databases, or that allows delivery across a network.

Loading pickled object arrays in npy files are allowed by default, however, loading pickled data can execute arbitrary code. Loading object arrays will no longer be possible if pickles are disallowed,

The 'allow\_pickle' parameter was introduced in numpy 1.10 version and the development team of numpy is now working on a fix.

“When using the 'numpy.load' function with the 'allow\_pickle' parameter users should default its value to 'False' if they are not sure if the data is safe,” Nan told [bleepingcomputer](https://www.bleepingcomputer.com/news/security/numpy-is-awaiting-fix-for-critical-remote-code-execution-bug/).

The maintainers of numpy also suggested to change the default value for the 'allow\_pickle' parameter to 'False' in numpy 1.17 version and display a warning when importing datasets from the internet, so that users can allow the action only if they trust the data.

**Linux-based distributions**

Numpy is typically available from the official package repositories of Linux-based distributions. SUSE security engineer Alexandros Toptsoglou stated that the issue impacts SUSE Linux Enterprise 15 and the SUSE Linux Enterprise 12 Service Pack 2.

Numpy is available for various Linux-based distributions via the RPM Package Manager. The maintainers of the numpy RPM package are open to adding a security alert in the near future so that it is present in Numpy version 1.17 if concerns are high enough to warrant such a move.

#### Scikit-learn

#### Scikit-learn is a simple and efficient tools for predictive data analysis that is accessible to everybody, and reusable in various context. It is built on Numpy, Scipy, and Matplotlib.

svm\_predict\_values in svm.cpp in Libsvm v324, as used in scikit-learn 0.23.2 and other products, allows attackers to cause a denial of service (segmentation fault) via a crafted model SVM (introduced via pickle, json, or any other model permanence standard) with a large value in the \_n\_support array. NOTE: the scikit-learn vendor's position is that the behavior can only occur if the library's API is violated by an application that changes a private attribute.