**1. How can learning from Study 1 enable us to identify and then exploit an opportunity in the Machine learning field? (500 words)**

The machine learning field has grown rapidly in recent years and more advancements have happened recently. Machine learning applications provide promising and reasonable solutions for many problems. There are so many opportunities in the machine learning field because of its potential and obvious outputs. This study helps us examine some of its major flaws and related issues in the software development phase. It is beneficial to know the common issues in the development because software development for machine learning requires both coding and data understanding skills. On the other hand, machine learning libraries are upgrading to solve many issues by resorting to robust functions but some practitioners don’t know about them. The developers or practitioners who are capable to handle the code smells can implement the models more precisely. In general, companies look for models which are efficient and faster than the others. Most people analyse the efficiency but do not look at the coding level problems. There is a high demand for models which are capable to address the problems with low code smells.

We can increase the efficiency of a model by taking precautions in the development phase, reduction of these issues will enable the model to stand in the industry. The code smells are the least concern while building a model, but the reduction of these smells will help the developers and companies to build more robust and structured models. Machine learning models work better while coding but due to the code smells and anti-patterns the model may become inefficient and less productive. It is very hard to find the code smells at the end of the model. This can be minimized if the developer knows about different kinds of code smells and solutions to reduce them. Code smells affects the model’s efficiency, maintainability, and complexity. The primary code smells are ‘misusing’ and ‘lack’ of knowledge about better functions. For example, the *pipeline()* function is useful for reducing the data leakage and this will help the issues downstream. Machine learning libraries provide less complex and easy-to-learn functions for solving complex problems. Exploring and implementing these functions increase the productivity of the model.

Using in-built pandas library functions will increase the efficiency of the code and it takes less time to execute. Pandas inbuilt functions will reduce the iterations and increase the efficiency. Efficient and proper memory usage plays a crucial role in machine learning models. This can be implemented by using different inbuilt functions like detach. There are many cases of misusing the functions and API’s. While implementing the functions the practitioners should take care of the parameters to avoid future problems. While handling the gradients, the developer should think about the order. Ignoring scaling will give inefficient results. Proper code smell handling can reduce the issues in the deployment phase. Controlling the generic code smells and API code smells will give many benefits like better performance, less error-prone, readable, robust, and efficient. So, learning these techniques is advantageous to both the practitioners and the companies.

**2. Write your thoughts about ”Opportunities implementing the process discussed in Study 1.”**

Machine learning code may give output according to the problem context, but we need to investigate the performance, efficiency, and understandability of the model. The model with code smells may give reasonable output with less efficiency. Many machine learning libraries provide easy solutions to these problems with in-built functions. Applying these solutions to a specific context will increase the performance and robustness of the model. Code smells are a concern to any group of people. It is a concern to data scientists, machine learning application developers, machine learning library developers, code analysis tool developers, and students. This is a continuous process of learning and implementing control methods to reach the goals. Implementing the new alternate functions can reduce the complexity and take less execution time. This process will help in increasing the efficiency, understanding, handling, and implementation of the machine learning model.

These solutions help in dealing with the Nan values and none values. These techniques help in the better implementation of scaling and gradients in the machine learning models. These methods are even useful with memory, parameters, scaling, randomness, mask, training, evaluation, data leaking, and misusing of the functions. The proper use of these methods will help in increasing performance and reduces the execution time of the code. The major advantage of these techniques is reduced future problems, manageable, easy to learn, and time reduction. They can be used in different types of domains and for different problems. These implementations will give appropriate outputs and efficient results.

When doing software development for machine learning we need to keep in mind the code smells and the future issues because of them. Many sources are providing promising solutions for these issues. Exploring these sources will help in implementing the code smell reduction. The major sources for suggestions are grey literature, paper mining, and library selection. Grey literature is getting information from the web. There are many good and efficient sources for suggesting alternatives to the functions with issues. There are also blogs written by the data practitioners which will help the students and new learners. Sources like stack overflow and GitHub will more information and they are easily accessible for learning. From paper mining and the citations from the papers, we can gather many code smells and solutions to reduce them. The papers will give deep knowledge about the topic. When coming to the library section, there are many libraries in machine learning to give solutions for different problems. Some libraries help more with a specific types of issues. For example, the Pandas library helps in dealing with iteration problems. Implementing these techniques on or before building a model will benefit the performance of the model.

This process is concentrated more on the code-level problems, and these will help the model in the long run. This process will help in building robust and efficient models with machine learning coding standards. This is a continuous process to sharpen the efficiency of the code and ease the problems in the production phase.