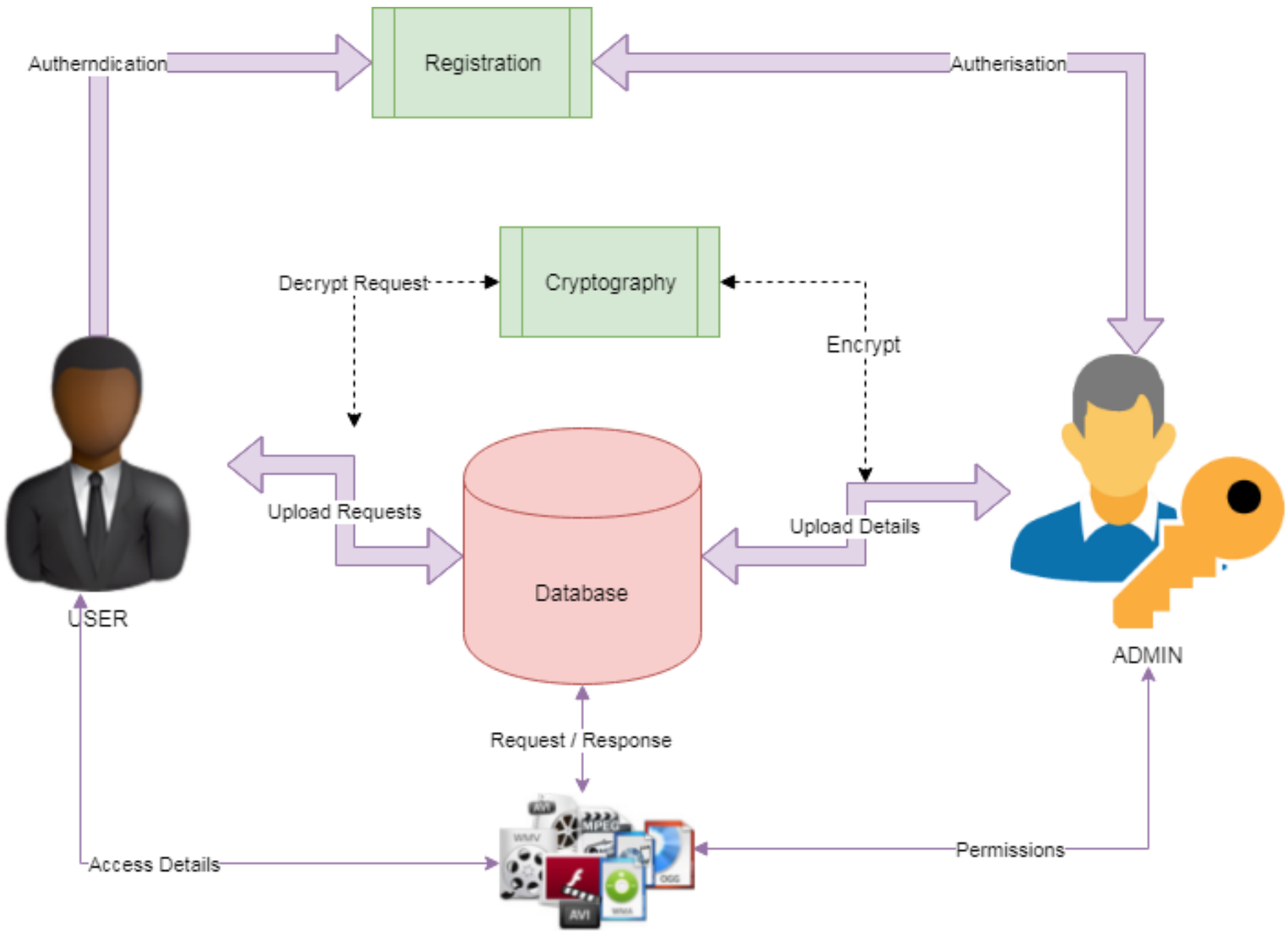
A BI-OBJECTIVE HYPER-HEURISTIC SUPPORT VECTOR MACHINES FOR BIG DATA CYBER-SECURITY

Cyber security in the context of big data is known to be a critical problem and presents a great challenge to the research community. Machine learning algorithms have been suggested as candidates for handling big data security problems. Among these algorithms, support vector machines (SVMs) have achieved remarkable success on various classification problems. However, to establish an effective SVM, the user needs to dene the proper SVM configuration in advance, which is a challenging task that requires expert knowledge and a large amount of manual effort for trial and error. In this paper, we formulate the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are considered as two conflicting objectives. We propose a novel hyper-heuristic framework for bi-objective optimization that is independent of the problem domain. This is the first time that a hyper-heuristic has been developed for this problem. The proposed hyper-heuristic framework consists of a high-level strategy and low-level heuristics. The high-level strategy uses the search performance to control the selection of which low-level heuristic should be used to generate a new SVM configuration. The low-level heuristics each use different rules to effectively explore the SVM configuration search space. To address bi-objective optimization, the proposed framework adaptively integrates the strengths of decomposition- and Pareto-based approaches to approximate the Pareto set of SVM configurations. The effectiveness of the proposed framework has been evaluated on two cyber security problems: Microsoft malware big data classification and anomaly intrusion detection. The obtained results demonstrate that the proposed framework is very effective, if not superior, compared with its counterparts and other algorithms.

Architecture



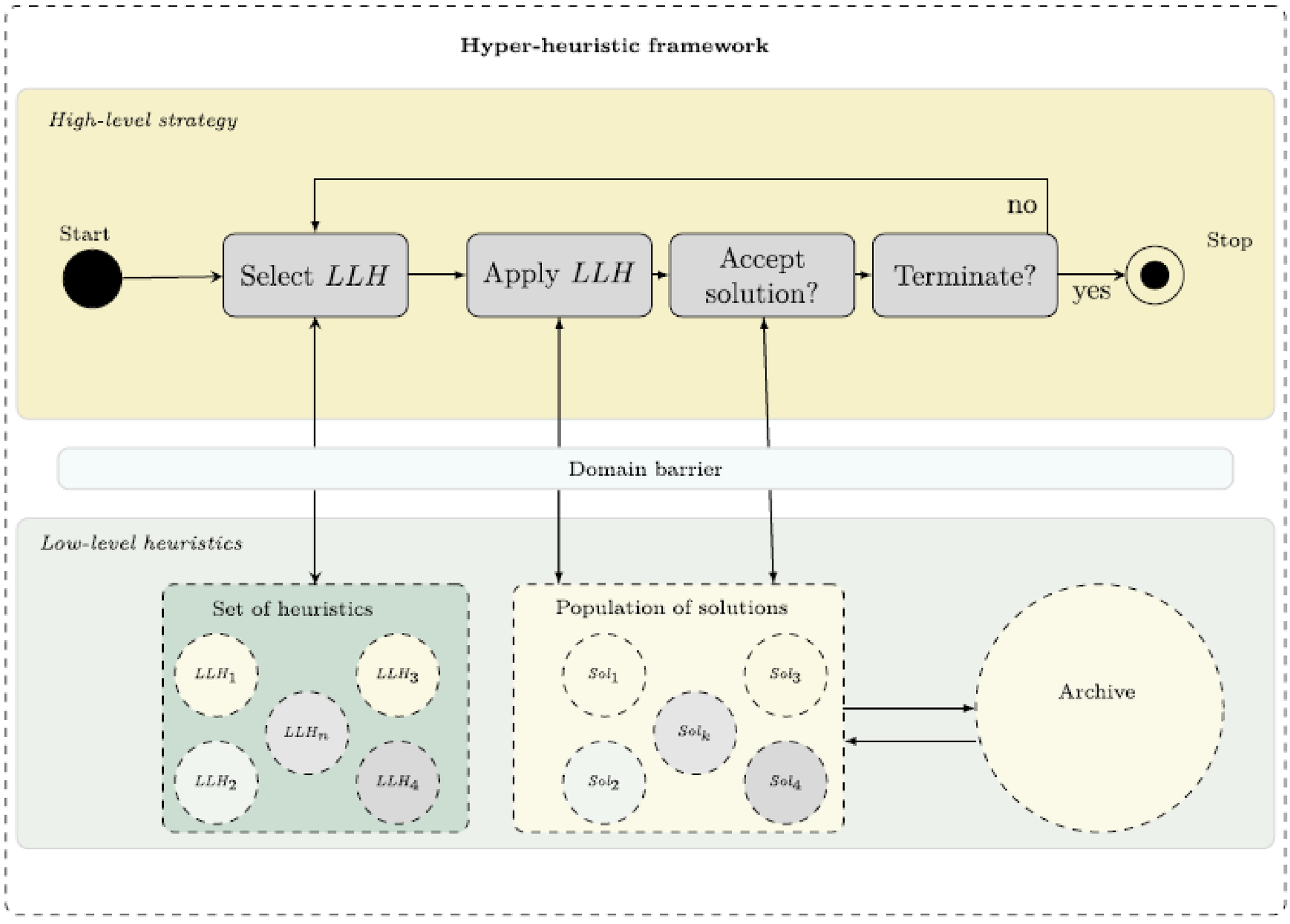
EXISITNG SYSTEM:

Cyber security in the context of big data is known to be a critical problem and presents a great challenge to the research community. Machine learning algorithms have been suggested as candidates for handling big data security problems. Among these algorithms, support vector machines (SVMs) have achieved remarkable success on various classification problems. However, to establish an effective SVM, the user needs to dene the proper SVM configuration in advance, which is a challenging task that requires expert knowledge and a large amount of manual effort for trial and error. In this paper, we formulate the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are considered as two conflicting objectives. We propose a novel hyper-heuristic framework for bi-objective optimization that is independent of the problem domain. This is the first time that a hyper-heuristic has been developed for this problem. The proposed hyper-heuristic framework consists of a high-level strategy and low-level heuristics. The high-level strategy uses the search performance to control the selection of which low-level heuristic should be used to generate a new SVM configuration. The low-level heuristics each use different rules to effectively explore the SVM configuration search space. To address bi-objective optimization, the proposed framework adaptively integrates the strengths of decomposition- and Pareto-based approaches to approximate the Pareto set of SVM configurations. The effectiveness of the proposed framework has been evaluated on two cyber security problems: Microsoft malware big data classification and anomaly intrusion detection. The obtained results demonstrate that the proposed framework is very effective, if not superior, compared with its counterparts and other algorithms.

Mail: **careers@aktechsolutions.in** **Phone:** **7777-976-476** **Web:** **http://aktechsolutions.in/**

PROPOSED SYSTEM:

The proposed hyper-heuristic framework for configuration selection is shown in Figure 2. It has two levels: the high- level strategy and the low-level heuristics. The high-level strategy operates on the heuristic space instead of the solution space. In each iteration, the high-level strategy selects a heuristic from the existing pool of low-level heuristics, applies it to the current solution to produce a new solution and then decides whether to accept the new solution. The low level heuristics constitute a set of problem-specific heuristics that operate directly on the solution space of a given problem. To address the bi-objective optimization problem, we propose a population-based hyper-heuristic framework that operates on a population of solutions and uses an archive to save the non-dominated solutions. The proposed framework combines the strengths of decomposition- and Pareto (dominance) - based approaches to effectively approximate the Pareto set of SVM configurations. Our idea is to combine the diversity ability of the decomposition approach with the convergence power of the dominance approach. The decomposition approach operates on the population of solutions, whereas the dominance approach uses the archive. The hyper heuristic framework generates a new population of solutions using the old population, the archive, or both the old population and the archive. This allows the search to achieve a proper balance between convergence and diversity. It should be noted that seeking good convergence involves minimizing the distances between the solutions and PF, whereas seeking high diversity involves maximizing the distribution of the solutions along PF. The main components of the proposed hyper-heuristic framework are discussed in the following subsections.



Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/

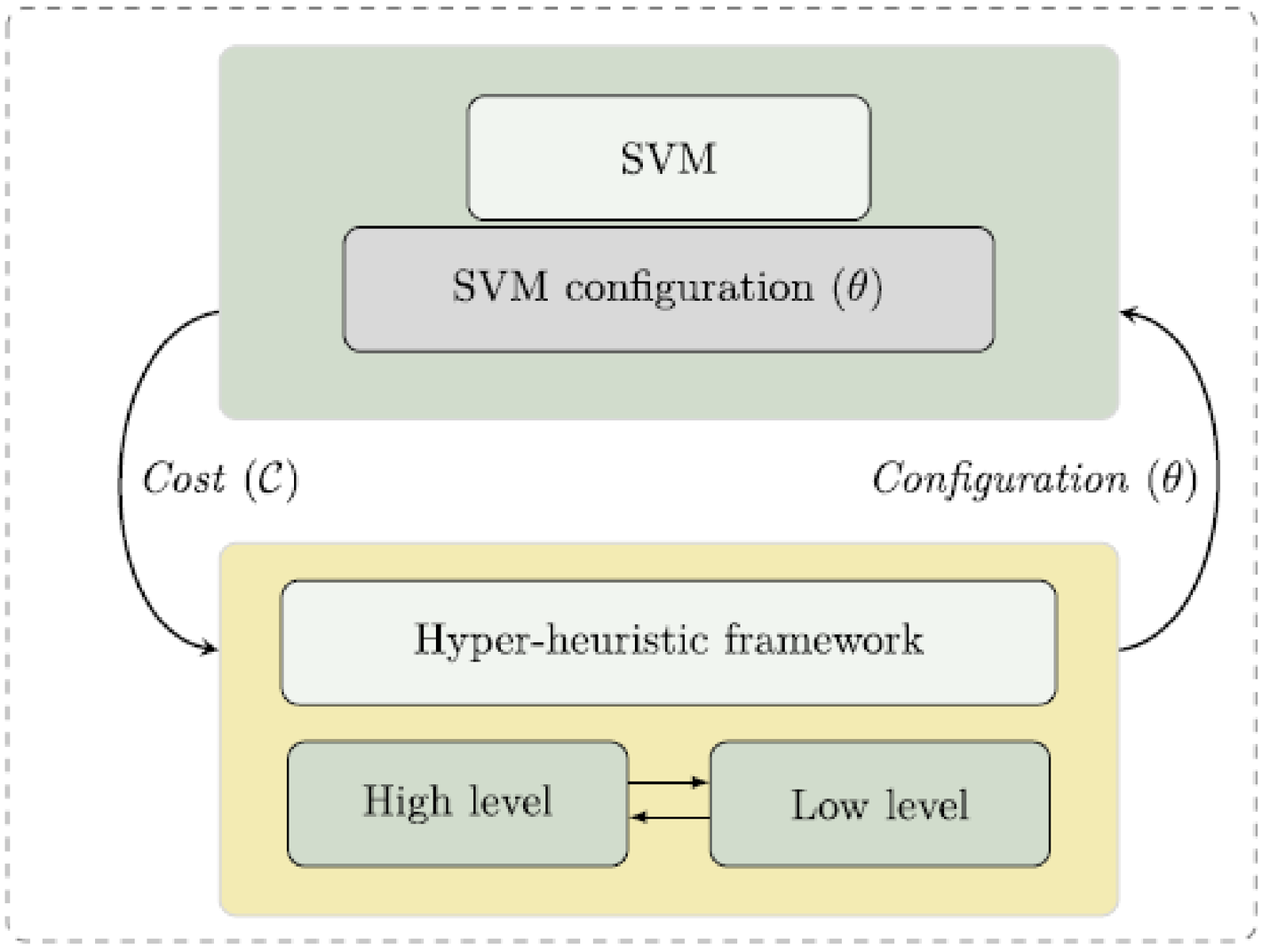


FIGURE 1. **The proposed methodology.**

MODULES:

The project is carried out based on the following modules listed:

1.Approved Users

In this system users are not allowed to access resources simply. User need verify their information's with admin. Admin are the authorized and trustworthy to the network. User need to send the request to administrator that they are interested to add the community. Admin views the user request and respond with the pass code to access users account through trusted sources like SSL (Gmail).

2.Security Steps and Upload

In this system users are not allowed to access resources simply. User need verify their information's with admin. Admin are the authorized and trustworthy to the network. User need to send the request to administrator that they are interested to add the community. Admin views the user request and respond with the pass code to access users account through trusted sources like SSL (Gmail).

3.Resource Access

The permissions to access the resource can be sent by users to admin. The requests have been updated by admin with the response to access the resource. Users can decrypt the resource and access the details. The important part is access the resource with the decryption. The passkey to access the details are limited. If the limit of wrong attempts over the threshold value means pass key expires.

4.Graphical Representation

This is graphical notation of the data given by the system. This phase of implementation will shows the effectiveness of the proposed system through pictorially in the order to better understand of proposed system.

ALGORITHM:

SUPPORT VECTOR MACHINE:

In machine learning, support vector machines (SVMs, also support vector networks) are learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier(although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

FUTURE WORK:

The current benchmark to compare our project is very small since to overcome this, our proposed framework has been tested on two benchmark cyber security problem instances: Microsoft malware big data classification and anomaly intrusion detection. The experimental results demonstrate the effectiveness and potential of the proposed framework in achieving competitive, if not superior, results compared with other algorithms. To do so some alteration to bring the exactness of the output can be achieved.

REQUIREMENT ANALYSIS

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

REQUIREMENT SPECIFICATION

Functional Requirements

Graphical User interface with the User.

Software Requirements

For developing the application the following are the Software Requirements:

1. **Python**
2. **Django**
3. **Mysql**
4. **Wampserver**

Operating Systems supported

1. **Windows 7**
2. **Windows XP**
3. **Windows 8**

Technologies and Languages used to Develop

1. **Python**

Debugger and Emulator

1. **Any Browser (Particularly Chrome)**

Hardware Requirements

For developing the application the following are the Hardware Requirements:

1. **Processor: Pentium IV or higher**
2. **RAM: 256 MB**
3. **Space on Hard Disk: minimum 512MB**

CONCLUSION

In this work, we proposed a hyper-heuristic SVM optimization framework for big data cyber security problems. We formulated the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are treated as two conflicting objectives. This bi-objective optimization problem can be solved using the proposed hyper-heuristic framework. The framework integrates the strengths of decomposition- and Pareto-based approaches to approximate the Pareto set of configurations.