UNIVERSITY OF GOTHENBURG

# Predicting Time Series Data collected from Software Measurements with Machine Learning Approaches

*Bachelor of Science Thesis in the Software Engineering and Management Programme*

HARRI PREENJA
MOHAMMAD ALI

University of Gothenburg
Chalmers University of Technology
Department of Computer Science and Engineering
Göteborg, Sweden, June 2016

**Predicting Time Series Data collected from Software Measurements with Machine Learning Approaches**

Harri Preenja
Mohammad Ali



Examiner: Miroslaw Staron

University of Gothenburg
Chalmers University of Technology
Department of Computer Science and Engineering
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Department of Computer Science and Engineering
Göteborg, Sweden June 2016

# Predicting Time Series Data collected from Software Measurements with Machine Learning Approaches

Harri Preenja
University of Gothenburg
Software Engineering and Management
Göteborg, Sweden
preenja.harri@gmail.com

Mohammad Ali
University of Gothenburg
Software Engineering and Management
Göteborg, Sweden
mohammad.ali@student.gu.se

***Abstract*— The objective of this paper is to highlight the implementation of machine learning forecasting approaches in software development. The concept of data mining has been used in different areas in the industry. There is an existing gap in the field of applying machine learning in the context of software measurements. This thesis will be conducted in two parts. Part 1, a systematic literature review to pinpoint the most recognised machine learning approaches. While part 2 will test the found approaches in an experimental environment to determine the most suitable machine learning approach for the collected data. The data was collected in a previous study through a collection of automotive software measurements.**

**Keywords— Data mining, time series, machine learning, forecasting, prediction.**

## 1. INTRODUCTION

Data mining and machine learning approaches are used more and more frequently, in multiple different fields of computer science [1]. As the gathering of data is huge in many companies, the management of sorting and retrieving valuable information from the data is the core element of current data handling. With various tools and new approaches, data can be analysed and provide future predictions, increase sales, prevent unwanted occasions among other things [1]. The different fields of data mining approaches that are studied in computers science more directed to software development are intrusion detection [2], effort estimation [3], and fault prediction [4], for example. Also, the specific field of machine learning for time series data is well studied [5].

The time series concept is described as a part of time in data where changes have occurred. This is then analysed by implementing different machine learning techniques as well as pattern recognition in the context of time series [6]. The result of forecasting can support decision making to reduce redundancy of code and predict certain events in software development. However these mentioned terms are ultimately interesting for future predictions which can give companies and others applying machine learning techniques a competitive edge. This thesis will examine how machine learning can be applied in a software development context by mining artefacts received from testing unit from an earlier case study [32]. The result can support decision making to create benchmarks for future growth in lines of code and complexity measurements, which may affect the prediction of certain events in software development.

### *1.1 Structure of article*

This article will be implemented in two parts, the first part will be a systematic literature review (SLR) and the second part will be an experiment. Firstly, from conducting the SLR, related machine learning approaches will be available to examine if patterns and forecasting is possible to implement on the time series received from earlier studies. This could provide a result of approaches that can be used for the specific time series data collected. Secondly, the experiment part (section 5) will be described, methodology and the result of the conducted experiment will be discussed and analysed. Finally, a conclusion (section 7) will be presented of the study to highlight key acknowledges and results relevant to the goal and research questions of this study.

## 2. PURPOSE OF STUDY AND RESEARCH QUESTIONS

There is a various amount of studies conducted in the field of data mining and the sub groups of the topic, such as machine learning, text mining, meta-learning, etc. A search in numerous scientific databases gives an overview of the interest and importance of the subject. However, the field is not fully explored, as new data and new approaches of recording data, gives the opportunity of future studies. Also, there is no previous research that is conducting the same study.

Even though the subject is highlighted from many aspects, software development has not embraced the possibility to create forecasts from mining software measurements to reduce future events that influence the process such as hours spent on software development. An opportunity has therefore opened to explore if the likelihood to extract valuable information from various sources that influences the code over time. To be able find which methods that exist and are mostly implemented we will have to examine the current literature to retrieve the best fit for collected data and evaluate the result. The purpose will be to see if what the theory and earlier experiments and case studies has found and if this will work as well when implemented on software measurements.

Machine learning and the different approaches in machine learning is highly important area for current and future data analysing [8]. For example, discovering patterns with the data mining approach can help to provide future decisions and prevent mistakes in many fields in society. Previous studies with a variety of algorithms have been developed in the field of different domains, such as machine learning, statistics, identifying patterns, etc [9]. The oldest one is the statistical algorithms, which also has a strict data distribution criteria,

compared to machine learning approaches. The machine learning approaches produce understandable patterns and have fewer restrictions regarding to data distribution. This is one of the reason that machine learning is found to be more popular [10]. As mentioned, there exist a variety of data mining applications, however, the purpose of the study is to identify machine learning approaches that are applicable to data in the context of software measurements.

### *2.1 Research Questions*

**RQ1** Which machine learning approaches are applicable to make predictions among time series data?

**RQ1.2** How do the applicable approaches perform regarding prediction accuracy?

### *2.3 Background and previous work*

The subject of data mining has been studied within different industries, certainly when it regards big data [1]. The concept of mining data has been around for a long time but applying machine learning techniques instead of statistical techniques on data are new to the field [11]. Chen et al. [11] displays an overview of the different existing concepts on data mining functionalities and the applicability of them in different regards such as time series, clustering and classification to mention a few. Prediction analysis on data and the comparison of different approaches applicable on time series data has been studied previously. However, the difficulty has been to forecast on no time-dependent variables [9], [10]. The use of prediction of future events has been implemented in occasions such as, energy management, telecommunications, pollution, bioinformatics and earthquakes [9].

In previous studies such as Malhotra [12], conducted a systematic review to identify which machine learning approaches were mostly used in finding software fault prediction. From the result he found that the three mostly mentioned and implemented machine learnings during his studies were the following; decision tree, neural network and support vector machine. Since our study concern software development it is an important factor to include these approaches in our context. Also, Malhotra mentions that from his findings most articles concerning machine learning in his context were published after 2007. Another important finding made was the comparison of traditional statistical methods and machine learning where he discovered that 19 of the 64 studies involved a comparison element. From the result Malhotra established that most machine learning approaches in these articles outperformed statistical linear regression [12].

Many published articles analyse the probability and the outcome of training machine learning approaches historically implemented in the industry. The purpose is usually to improve the forecasting accuracy by training of the algorithm or using another training algorithm before the forecasting is implemented. The machine learning approaches that are mentioned in those studies is support vector machine or artificial neural network [13], [14].

To summarise this section, the subject of machine learning forecasting has been studied considerably well. Most of the studies concerning time-series forecasting focus on the economic, electricity and health field. The research is focusing on developing and improving forecasting approaches to achieve improved accuracy. However, the comparison and applicability of machine learning approaches on software development data with the goal of forecasting, has not been examined to answer the purpose of this study.

## 3. METHODOLOGY

The methodology process (figure 1) of this research will be conducted in two steps, starting with a systematic literature review (SLR), then followed by an experiment. The aim of the SLR is to understand what current research have been discovered on machine learning approaches and time series forecasting. This will propose the relevant approaches to implement on time series data. The aim of the experiment is to implement and compare the approaches retrieved from the SLR. The two steps of methodology will be described more in detail in the coming sections.

### *3.1 Literature Review*

#### *Search Strategy*

The search strategy for the SLR will be described in this sub section, this is to clarify the steps implemented to reach the results of the SLR. Firstly, the aim was to create a proper search string that could be implemented in the databases for published articles. Several trial searches were done in order to receive a final search string, namely to determine that the keywords that were chosen appeared in the required fields in the database search result. In order to find previous conducted systematic literature reviews of the subject, a search for published systematic-reviews was processed. To implement a proper search, a method was used known as subject heading. The purpose of using the method is to provide previous published articles, such as already published systematic literature reviews. The strategy of using subject heading search for this study is to find previous published SLRs.

The systematic review search strings has been created for the purpose to filtrate the nonessential articles in the result of the databases. The search strings are based on the keywords that are identified in the research question and contain the necessary inclusion keywords. The combination of the words in the search strings aim to retrieve the best result for the ambition of the search. However, synonyms of the keywords are used to prevent missing relevant articles for the literature review. The search indicated that machine learning/data mining/artificial intelligence and forecasting, prediction could appear as synonyms in relevant articles.

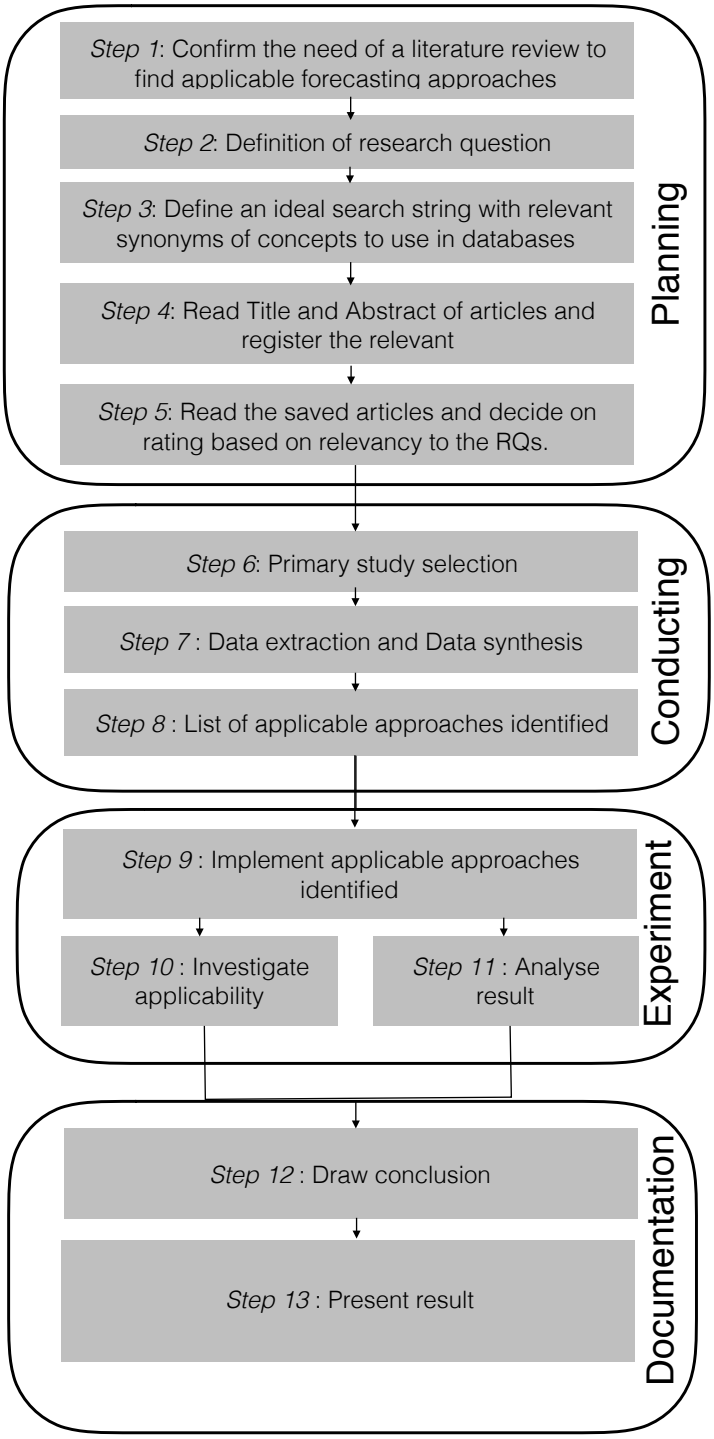



Figure 1. Methodology process

For our search criteria there are some important keywords that has to be included somewhere in the range of title, abstract and subject. The importance here is that some terms such as data mining and machine learning are vast and can occur in many contexts. Two of the selected database engines namely, Inspec and Scopus (figure below), provides suggestions for synonyms of the keywords, which is helpful to determine the final creation of the search string. Another benefit is that the search engine provides available synonyms in the topic, this appears as a recommendation and it is built in. As Kitchenham [15] states the use of electronic databases is a first step at initiating the first searches of the primary studies for the systematic review.

**SCOPUS**

( TITLE-ABS-KEY ( "machine learning" ) OR TITLE-ABS-KEY ( "data mining" ) OR TITLE-ABS-KEY ( "artificial intelligence" ) AND TITLE-ABS-KEY ( "forecasting" ) OR TITLE-ABS-KEY ( "prediction" ) OR TITLE-ABS-KEY ( "pattern recognition" ) AND TITLE-ABS-KEY ( "time series" ) ) AND PUBYEAR > 1999 AND ( LIMIT-TO ( LANGUAGE , "English" ) )

For the search string used in the different databases, a combination of booleans are used to combine the search terms for filtering. The search string contains the boolean of AND, NOT and OR. For complex searches in the subject headings, the use of the booleans with synonyms helped to provide the most of the search.

The strategy implementations mentioned above, helped the search strategy to obtain a complex search to filtrate the unwanted articles. In this way the search string was optimised to produce relevant results. To narrow our search further, Hilbert & Lopez [16] explains how the majority of information has been stored in digital format since the early 2000s. From these findings we have decided to limit our search string to begin at the year 2000 because of the escalating use and improvements of machine learning since then [9]. The publication year interval was filtered, this was for reports published the year 2000 to early 2016 since the literature review was conducted in march 2016. The search strategy implemented resulted in articles related to the research question and the purpose of the study.

*Selection Criteria*

The selection criteria will be specified to find the relevant literatures due to type of the received data, recent published articles, languages, etc. The selections criteria specification will be discussed in detail below.

Regarding the data for this study; it is time series data collected in previous studies from software measurements. As a result of the data type, time series data, approaches and articles focusing mainly on big data will be excluded from the selection of literature. The decision to exclude big data articles will result in an inclusion criteria to find literature that is focusing or analysing different machine learning approaches used on time series data. Due to the fact that the data set used in this study is not regarded as a big data problem, and also, as mentioned previously, articles focusing mainly on big data will be excluded from the selection of literature. As we intend to find methods applicable in the computer science domain, we were looking for approaches that have proven to work earlier in industry or artificial environment implemented on time series.

There will also be a limit in the exclusion on basis of the date of release since the field is quite researched with approaches being developed early because of the statistic value. Papers describing definitions, technical, and theoretical aspects of machine learning approaches can therefore be excluded. In our context, papers covering application and implementation are relevant.

*Exclusion*

- Non-fulltext
- Non-English
- Duplicates
- Training algorithms related articles
- Non computer science
- No articles with big data analysing

*Inclusion*

- Only journal and papers from academic databases
- Keywords must be included in title, abstract or keyword section
- Subject area must concern machine learning approaches
- Articles with approaches implemented on time series

Table 1: Amount of articles retrieved

| No | Source | Number of found articles |
|---|---|---|
| 1 | SCOPUS | 3265 |
| 2 | Inspec - Engineering Village | 2929 |
| 3 | Science Direct | 1437 |
| 4 | IEEE explorer | 5636 |
| 5 | ACM Digital Library | 119 |
| **Total** | | **13386** |

*Data extraction*

From the found articles (table 1) that resulted from our search query, these criteria were extracted:

- The source of the article, i.e., the database used.
- The authors and full reference
- Addressed research questions
- Abstract
- Main topic
- Machine learning definitions
- Data collection strategy

The information that was extracted through the criteria above should provide sufficient amount of data to provide solutions for our research questions. In our selection of articles we reviewed them in ascending order sorted by relevance which meant that the database search engine algorithm was trusted as a starting point in the selection process.

In a systematic manner the data extraction procedure started with the selection criteria that was applied on the title and abstract, and also if necessary, on the introduction and conclusion of the publication to decide if it should be saved and assessed or not. The process of going through the found material was evaluated simultaneously by the researchers. This gave room for discussion if needed ambiguity was present.

Due to time limitation a full systematic review goes beyond the scope of this study and therefore a desire to conclude the searching for material at a certain stage. From the earlier pilot searches made when creating the final search string we noticed a recurring problem of studies that were of low sufficiency for our work after the 40 first articles. With the sophisticated algorithms that most scientific databases uses when filtering relevance we found a way to limit our literature review. Therefore a strategy to find the best quality we decided that if 0 articles were found within 20 articles twice i.e. no satisfactory material for 40 read articles.

The conclusion that nothing more relevant for our context would be found and therefore continuing to exploring the next database. The same model was earlier used by Unterkalmeister et al. [17] as they evaluated papers during their pilot and concluding that nothing more could be found. From this procedure we went through 520 articles and found the same pattern that nothing of our interest kept repeating itself through all five databases that was used. The result from our review concluded with 61 articles that would be used for further reading in our iterative process.

*Data synthesis*

The derived data from the selected studies according to the extraction criteria mentioned above will be examined and compared with regard to the research questions. The synthesis will consist of two tables which has a quantitative structure listing the findings which pass the criterion when applying the search string. The other will have a descriptive nature to complement and give a subjective view of how the literature interpreted our research question and if it was sufficient enough.

*Quality assessment*

When going through the main articles for the study the quality assessment gauge were the following:

1. Is the goal of the research satisfactory explained?
2. Is the given approach or proposal explained enough?
3. Are the results as well as interesting findings of the research stated?
4. Are the review's inclusion and exclusion criteria described and appropriate?

By applying these questions we can assure that the extracted data can be examined further and create our framework for the quality assessment. Furthermore the framework can provide help to decide the appropriateness for analysing the data. The first four questions protects any unwanted data by giving a clear content of the purpose in the topic and research. The last question highlights the importance of removing articles that might be affected by bias result and therefore contributing to loss of validity.

### *3.2 Experiment overview*

The experiment part of this article will be conducted after the SLR result. The results found from the SLR, providing evaluation of machine learning approaches from previous studies will be a leading point to support the implementation of the experiment in this study. To precise how the implementation of the experiment will be done, we conclude from the article *A Holistic Overview of Software Engineering Research Strategies* [18] that the study has elements of an experimental simulation. The description of different approaches of research strategies, determines the strategy of this study to Experimental Simulation. The Experimental Simulation is when the activities used in the research has been collected from a natural process. However, the simulation part is a "place" that has been created artificially and that is also where the behaviour is observed [18].

From interpreting Stol & Fitzgerald [18] the nature of this study will be an experimental simulation approach. Since, some of the limits and regulations are set such as the data other proposals could not be considered. Namely the data that is given cannot be changed or manipulated in any way but instead it can be used and applied indifferent settings. Also since the data is taken from an industry research it is more aimed through an obtrusive setting instead of building on theoretical background and frameworks studies [18]. However, the detailed description of the experiment implementation will be presented in the Experiment section (5).

## 4. RESULT LITERATURE REVIEW

### *4.1 Data Extraction*

Conducting the study selection from our research strategy, 61 publications were found that were accepted for the data extraction. The process was made iteratively by ranking each paper relevant to the research questions and related inclusion criteria. We extracted the following criteria mentioned in table 2, which based the relevance scored and a discussion was made after each paper. This procedure was to extract the mentioned criteria without interpreting the content of the study too much. Also looking for the context in regard to industry related studies or with data sets from real world domains counter to artificial data. From this a list of 15 most highly relative articles was derived (table 3).

Table 2: Extracted Properties

| ID | Property | Research question(s) |
|---|---|---|
| P1 | Approaches / Methods | RQ1 |
| P2 | Case study / Industry | Overview of Studies |
| P3 | Comparison of approaches | RQ1 |

Table 3: Selected articles

| ID | Source | Authors | Year | Title | Topic Area | Approache category |
|---|---|---|---|---|---|---|
| S1 | ScD | Bose, Indranil<br>Mahapatra, Radha K. | 2001 | Business data mining — a machine learning perspective | Business decision support | NN, GA |
| S2 | EnV | Carpinteiro, O. A. S., Leite, J. P. R. R., Pinheiro, C. A. M., & Lima | 2012 | Forecasting models for prediction in time series | Artificiell intelligence | SVM, ANN |
| S3 | EnV | Choi, Tsan Ming., Hui, Chi Leung. Yu, Yong | 2011 | Intelligent time series fast forecasting for fashion sales: A research agenda | Sales forecasting | ANN, ARIMA, ELM |
| S4 | IEEE | Crone, S.F.<br>Lessmann, S.<br>Pietsch, S. | 2006 | Forecasting with Computational Intelligence - An Evaluation of Support Vector Regression and Artificial Neural Networks for Time Series Prediction | Sales machine learning | SVR, ANN |
| S5 | IEEE | Garcia, M.P.<br>Kirschen, D.S. | 2004 | Forecasting system imbalance volumes in competitive electricity markets | Electricity markets | NN |
| S6 | ScD | Gerdes, M. | 2013 | Decision trees and genetic algorithms for condition monitoring forecasting of aircraft air conditioning | Genetic Algorithms | Decision Tree |
| S7 | ScD | Güiza, Fabian., Ramon, Jan Bruynooghe, Maurice | 2009 | Machine learning techniques to examine large patient databases | Health care | SVM, ANN |
| S8 | ScD | Kattan, Ahmed., Fatima, Shaheen<br>Arif, Muhammad | 2015 | Time-series event-based prediction: An unsupervised learning framework based on genetic programming | Artificially generated data | GP |
| S9 | EnV | Lindsay, David<br>Cox, S | 2005 | Effective probability forecasting for time series data using standard machine learning techniques | Machine learning performance | SVM, HMM |
| S10 | ScD | Lykourentzou, Ioanna., Giannoukos, Ioannis. Nikolopoulos, Vassilis. Mpardis, George. Loumos, Vassili. | 2009 | Dropout prediction in e-learning courses through the combination of machine learning techniques | E-Learning | NN, SVM, ARTMAP |
| S11 | ACM | Martínez-Álvarez, Francisco Troncoso, Alicia Asencio-Cortés, Gualberto Riquelme, José | 2015 | A Survey on Data Mining Techniques Applied to Electricity-Related Time Series Forecasting | Electricity markets | SVM, ARIMA, ANN, etc |
| S12 | ScD | Radzuan, Nabilah Filzah Mohd Othman, Zalinda<br>Bakar, Azuraliza Abu | 2013 | Uncertain Time Series in Weather Prediction | Weather forecasting | NN, |
| S13 | Scopus | Sapankevych, Nicholas<br>Sankar, Ravi | 2009 | Time Series Prediction Using Support Vector Machines: A Survey | Machine learning approaches | ANN, SVR, SVM |
| S14 | Scopus | Vanajakshi, Lelitha<br>Rilett, Laurence R. | 2007 | Support vector machine technique for the short term prediction of travel time | Transportation | ANN, SVM |
| S15 | IEEE | Yoo, P.D.<br>Kim, M.H.<br>Jan, Tony | 2007 | Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation | Stock market | NN, SVM |

*Note*: Artificial Neural Network (ANN), Neural Network (NN), Support vector machine(SVM), Support vector regression (SVR), Gaussian process (GP), Extreme learning machine (ELM)

*Approaches*

The various machine learning approaches that exist have been implemented and evaluated in different settings in an amount of studies. According to the result of the 61 studies that were assessed, the practice of the approaches implemented in machine learning can be categorised into linear and non-linear. Depending on the case or the order of the data, another attribute of the approaches could be supervised or unsupervised learning of the approach. Further, the order of the data and the learning time has to be considered before an approach can be proposed as superior to another. However, there is a few of the approaches that generally are recommended to be applicable to extensive amount of data types with satisfying accuracy result of forecasting.

Machine learning as the name indicates has different approaches for each algorithm on how to learn from the target data and from that experience being able to predict future outcomes. The chosen method is dependent on the knowledge of the data, if the data is clean, namely structured and labeled which gives the miner and insight in what the outcome should be. Otherwise it might be more interesting to look for hidden patterns and look at the data as a whole and through the algorithms detect similarities which can cluster and then find interesting groupings. Bose et al. [10] defines it as "Machine learning is the study of computational methods to automate the process of knowledge acquisition from examples." The study of machine learning evolved to eliminate laborious and expensive knowledge engineering process involved in developing knowledge based systems [10].

The approaches/techniques implemented in machine learning is mainly categorised in two data learning attributes, such as supervised and unsupervised learning [8]. The supervised learning is more concerned with teaching the algorithm to detect outcomes and through analysing the result detecting if some outcome is missed and therefore through supervision, learning the algorithm to in the future detect the missed case. The aim is to create a model which can predict the value of the target variable for new unknown data. Guiza et al. [8] states that if the target variable is nominal then the prediction task is known as classification.

However unsupervised learning indicates that the modelling is handled with an unknown target variable, namely descriptive [8]. The purpose of the process is to discover interesting consistency, that describes similar subgroups and clustering them together. For instance even though the model is created with unsupervised learning is mad for a sole purpose, other hidden patterns that were not intended findings may become apparent thanks to the nature of unsupervised learning.

*Linear and Non-linear*

The time series forecasting is another parameter that is present in the case of machine learning. Time series is defined as a sequence of data points that are measured in controlled time intervals [18]. The time-series data gives a natural sequence of events since the point of time can be the same compared to other data analysis with mixed time intervals.

Forecasting in time-series can be proceeded in several ways, Kattan et al. [18] mentions two different forecasting methods that are either based on linear functions or non-linear. The linear forecasting tries to create a model for time series behaviour methods based on linear functions [18], [9]. Distinctions between them offer different abilities, e.g., non-linear methods, whose primary advantage is that there is no need to know the input data distribution.

*Time series forecasting*

The concept of time-series in machine learning is a measure of sequence points at a certain point. Contrary to common data analysis problem, time series provides a distinct data analysis due to the nature temporal order of time series observation [6]. Time series shows an understandable representation which strengthens the use of it and also makes it easier for comparison if the time intervals represent the same time durations during observations.

Common features that can be found in time series are trend and seasonality [19], [9]. As Wang and Smith-Miles stated, it is natural for a time series to be characterised by its rate of trend or seasonality. Furthermore they discovered that that once a time series is measured in regard to their type i.e. seasonal or trending it can be detrended or deseasonalized which can make it applicable for noise or chaos implementation. To distinct between these established concepts, one must know what differs between them. For trending time series is the general movement that the variable displays during the examination period [9]. A more concrete term for trend is when there exists a long-term change in the mean level. There is also a common perception that in the research field where they refer to it as the long-term movement that the time series presents. Trends can show different various profiles being linear, exponential or parabolic [19].

Seasonal patterns on the other hand exhibits seasonal factors in time series, which can be of periodic basis such as day of the week or month of the year that affect the time series. The definition of seasonality regarding time series is when a pattern is repetitive over fixed intervals of time [19]. Seasonality might emerge from several aspects such as weather conditions, economical cycles or holidays.

*4.2 Data Synthesis*

*Most popular linear approaches*

As discussed in Martinez et al., [9], the existing popular techniques implemented as linear approaches are the following:

- AR - Autoregressive
- VAR - Vector autoregressive models
- MA - Moving Average
- ARMA - AR and MA
- ARIMA - Autoregressive integrated moving average
- ARCH - Autoregressive conditional heteroskedastic
- GARCH - Generalised autoregressive conditional heteroskedastic

The traditional statistical models that are used are generally towards economics for time series prediction [20].The models are based on the same methodology introduced by Box and Jenkins [21]. The linear techniques limitation is that they presume that the time series are linear (the values are normally distributed); prediction of values for a certain variables is based on the previous one [9], [22].

*Most popular nonlinear approaches*

*Support Vector Machine*

Support Vector Machine (SVM) is an approach implemented and evaluated in many machine learning studies [24]. The approach is based on finding vectors (points) in the data as support vectors to forecast [23, LYKO]. An advantage of SVM is that the accuracy is not affected when it identifies the support vector in the data. When the approaches selects a support vector, the training is independent from the amount of data, this is the SVM advantage of the performance is not concerned of the amount of the training data [24], [8].

*Artificial Neural Network*

Artificial Neural Network (ANN) is a machine learning approach inspired by the human neural networks. The approach is based on processing the relation between the nodes during the learning of forecasting. The input node or the observed quantities is used for the prediction, this is also a step of the learning process to find the error and correct the output from previous nodes [8]. That is to say, ANNs performance and accuracy is relying on the amount of the data; greater historic data, a more accurate forecasting with the ANN approach [24]. Other implementation of Neural Network is Extreme Learning Machine (ELM) and Self-Organizing Map (SOM). The ELM approach uses a faster training method the classical neural network. The SOM on the other hand does not need to be supervised when implemented in difference to classical Neural Network [9].

*Decision Tree*

Decision tree is a method used in machine learning and it is built of binary trees (if-then else). There is also more complex types of decision tree using more than two branches using switch function. To implement decision tree for prediction, training has to be implemented (e.g., ID3 algorithm) [25], [26].

*Combination and training of approaches*

A smaller number of articles discusses the implementation of a combination of approaches together to achieve the best result of forecasting or accuracy. The concept of combining approaches to increase predictive performance is known as ensemble models [9]. It has recently received attention from the research community and it is very interesting due to increased performance and solving classification problems. The concept has been implemented in prediction of energy time series by using data clustering learning model such as Bayesian clustering, before implement the SVM approach [9].

Another technique that could be implemented as a learning approach before implementing SVM or ANN, is the Particle Swarm Optimisation (PSO). The method is used to train the approach such as ANN or SVM. However, the PSO is a good method implemented on certain time series data. The method is performing unsupervised and can discover relations in the data set, this is beneficial to decrease the error rate of the approach [27].

*4.3 Results*

Table 4: Popular machine learning approaches

| Approach | | | |
|---|---|---|---|
| *Linear* | Training Data | Weakness(–) / Strength(+) | Applied in industry context |
| AR | Supervised | - Restrictive conditions of the data structure [28] | Statistical |
| VAR | Supervised | - Restrictive conditions of the data structure [28] | Statistical |
| MA | Supervised | - Restrictive conditions of the data structure [28] | Statistical |
| ARMA | Supervised | - Restrictive conditions of the data structure [28] | Electricity [9] |
| ARIMA | Supervised | + Good when aim is to find structures on basic stationary data structure [22] – Are limited to predicting values for one variable based on its pervious values [22] | Statistical Electricity [9] |
| ARCH / GARCH | Supervised | - Restrictive conditions of the data structure [28] | Electricity [9] |
| *Non-linear* | | | |

| Approach | | | |
|---|---|---|---|
| SVM | Non Supervised | + Accuracy is not affected when it identifies the support vector in the data [24]<br>+ Good when data is limited/short-term [24]<br>– Long training process [8], [29] | Electricity, Travel time, Medical, E-learning [23], [9], [24], [8] |
| ANN | Supervised and unsupervised | + When more data available to learn from better accuracy then SVM [24]<br>– Long training process [8], [29], | Electricity, Travel time, Fashion Sales marketing, Artificial data [9], [24], [28], [26], [30] |
| ELM | Unsupervised | + Fast learning | Electricity , Agriculture, Fashion sales marketing [9], [28] |

Source: [7], [9], [22], [24], [28], [29]
Note: + advantages, – disadvantages

As mentioned in the previous data synthesis the decision of machine learning towards non-linear is going to be used and evaluated to later be implemented in an experimental scenario. Previous studies as Guiza et al. [8] used the same motivation when examining large patient databases for clustering and unknown data the machine learning methods do not require the same amount of detailed information about the data when being applied [8]. Also previous work showed that ANN, a popular neural network, outperformed linear regression 10 out of 28 times, while linear only outperformed ANN four times. Another factor behind our decision is that non-linear approaches are proven to handle noisy data in a superior way in contrast to linear approaches [10]. This argues for the reasoning to experiment with the non-linear approaches on the collected data to compare the results of the approaches. Also, Crone [30] motivates in his study that NN outperforms statical approaches which also was stated in previous studies.

From the retrieved information and the summarisation in table 4, RQ1 is now finalised. The given approaches that will be used for this experiment is ANN and SVM. The RQ1 purpose is to find the most implemented approaches within time series data. The table 4 extracted from the literature argues for the most implemented approaches with relation to the context of the study. The foundation of the experiment will be based on the result of SLR in the next section.

## 5. Experiment

The nature for the experiment will follow Basili et al. [31] framework for the reasoning behind why the experiment will be carried out. Basili et al. [31] argues that experimentation is carried out in order to get a deeper understanding of predicting, improving and understand products as well as processes regarding software development. In addition experimental proceedings involves the activity of a hypothesis and test process. Finally Basili et al. [31] states that the main reason for experimental studies is to enhance the knowledge and understanding for the given context. Following their structure we have identified which areas we want to achieve for evaluating the result and why [31]. Firstly a clear definition and the problem is stated:

- **Motivation:** To asses different prediction approaches might provide different results depending on the context. The motivation for this experiment is to understand the differences of applicable machine in the field.
- **Object:** Process and model of collecting data measurements from development by applying support vector machine and artificial neural network on the data set.
- **Purpose**: To find out if there is a possibility to predict future outcomes of software complexity from several aspects. Evaluate the accuracy of the approaches to determine which one suits the context better.
- **Perspective**: The stakeholders that might gain from findings in predicting data can be developers and project managers. The argument for the developers are because they are in direct contact with code, and therefore know when to react to warning signs. The project manager will easier track the process if it is possible to forecast the code simulation.
- **Scope**: Single project. A case study was completed were that data collection was made that will be examined in this experiment.

The aspect we are investigating is the time series of software development measurements consisting of four parts, namely size i.e. lines of code. Secondly the other size measurement is amount of present modules or packages namely block count. The two complexity measurements coupling and cohesion which show how the different modules are connected. What the industry is interested in is predicting the future and therefore avoid mistakes and learn from previous data.

As the main purpose for the experiment part is to analyse the two approaches from the LR result, there is a need to frame the hypothesis for the experiment that can clarify the accuracy level of the two applicable approaches. The comparison of the accuracy level of the implemented approaches will also determine which one is the most applicable approach with best accuracy to the collected data from the industry. The following hypothesis will be tested:

**$H_0$** The machine learning approaches SVM and ANN gives the same forecasting accuracy on time series data.

**$H_1$** A certain approach, SVM or ANN gives higher accuracy in forecasting.

*5.1 Data and Methodology of experiment*

The section will describe the background of the data and how it was collected. Also, the implementation of the experiment methodology will be described to clarify how the result was retrieved.

*Data*

This study will use data sets that was previously collected by Schroeder et al. [32]. They performed software measurements on simulation models in the automotive industry. Measurements were performed on all past software revisions over time, creating time series of measurement data. Four metrics were used. They are the main data source for this study. The metrics size which will be represented by lines of code (LOC). This will be the main aim for our accuracy comparison since the increasing code evolves is of high interest from an industry context. Another attribute that illustrates the size is the block count which display each block in the model, including internal ones in subsystems. Complexity will consist of a structural complexity attribute (SC) and data complexity (DC). The SC gives the number of the coupling and the DC shows the cohesion. However, the data complexity measures the amount of work one block in the model has to perform [32].

These four attributes all have correlations between them, they represent the same data set but from different perspectives. Each file represents a component in the system so in total all files consist of several packages. We were given a total of 71 files, from these they were all structured in the same way consisting of 5 columns which were mentioned before. From these 71 files, 10 were removed because of being empty or due to missing data and consisting of too small data sets for prediction. The data does not have any timestamps since the revisions for when it was collected varies. The value for the revision is represented as a numeric id which has an ascending order which shows that the greater the number the later the revision is for the component. This factor helps to understand how the behaviour changes from a time aspect even though the exact time range is unknown. All the data columns consist of numeric values and can therefore not be changed in any way. The numeric intervals for the data contains of ascending, descending trends with many leaps in values. There was files that had missing values and therefore since it was imperative for our prediction to use the attributes of the missing values had to be removed.

Table 5: Example of file structure for accuracy calculation

| Model | Revision | SVM(SMOreg) forecasting result LOC | ANN(MLP) forecasting result of LOC | Actual LOC | Actual LOC mean / SVM LOC forecast mean | Actual LOC mean / ANN LOC forecast mean |
|---|---|---|---|---|---|---|
| *M1* | *1* | *10350* | *15306* | *11500* | | |
| | *2* | *14145* | *8247* | *12630* | | |
| | *3* | . | . | . | | |
| | *Mean* | *12247* | *11776* | *12065* | *0,985 (98,5%)* | *1,024 (97,5%)* |
| *M2* | . | . | . | | . | . |
| . | | | | | | |
| . | | | | | | |

Note: *The data in table is example data and do not represent the real data from the revisions.*

*Methodology*

The sorting and filtering of the received data was handled in several steps. During the process of data handling, there was a need to edit the data and the structure to avoid poor forecasting due to data structure or insufficient values. The structure was sorted from highest value in revision to lowest. To make a prediction for upcoming data, the forecasting is based on previous revision data. To receive the optimal setting for the experiment, a pre-test was made with the two retrieved machine learning approaches on one model. The purpose was to identify the combination of attributes that was most suitable for prediction. However, the data also had lines with missing values in the revision, those were deleted to avoid poor

forecasting generated because of the missing values. The handling and sorting of the files and was conducted as follows:

1. Firstly, a review of the data content was conducted to avoid corrupt data such as rows or columns with partly missing or insufficient information. The data was cleared of insufficient data information to avoid poor prediction due to corrupt data. Secondly, the rows were sorted to ascending revision, the original file was sorted with the highest revisions descending. The purpose of sorting the revisions into an ascending structure is to use the previous revision to predict the result of the coming revisions, and structure in ascending order is necessary to have a correct prediction in time index.

2. For all further analyses the WEKA (Waikato Environment for Knowledge Analysis) machine learning tool is used. The WEKA machine learning tool is a popular tool used for machine learning. WEKA is free to use and the learning curve is steep for the users. Also, it has high portability to implement many of the popular machine learning algorithms, this and many other features makes WEKA a popular suite of machine learning.

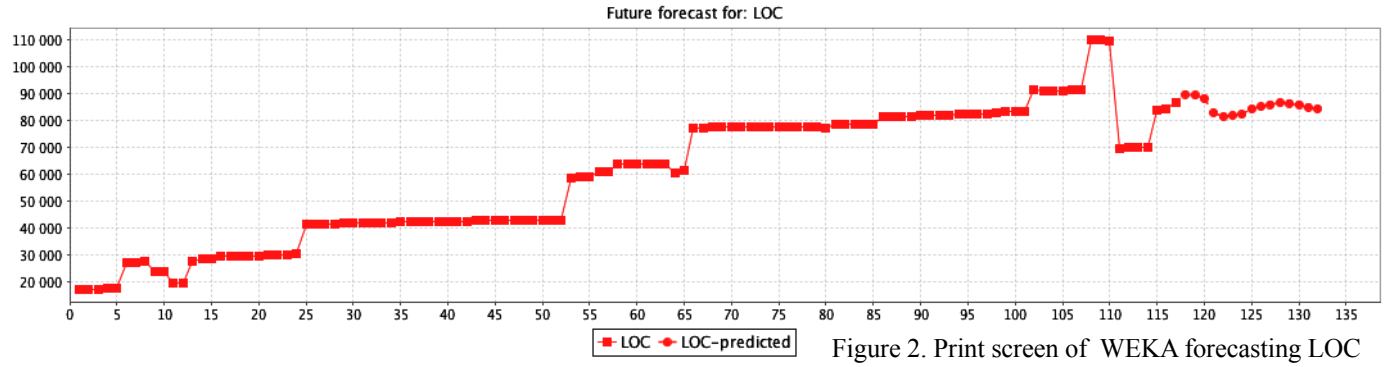



Figure 2. Print screen of WEKA forecasting LOC

The machine learning approaches that is used in WEKA for this experiment is Multilayer perceptron (MLP) and Sequential minimal optimisation (SMOreg). The MLP is an artificial neural network model and the SMOreg is a support vector machine for forecasting. Another condition in WEKA is that it expects input files to be in .csv format. However, the files were received as .txt format, the format type could therefore not be imported to WEKA. In order to handle the data files, they were imported to Microsoft Excel to conduct the sorting (ascending), the files was saved in a .xlsx format. Finally, the files were converted to .csv that is an acknowledged file format in WEKA.

3. To compare forecasted data to actual result it was necessary to move 25% of the original data to another table. This made it possible to forecast from 75% of data up to 100% then compare the forecasted data to the actual data in the new table. Example: If there is 100 rows of data, a cut of 25 of the rows is made. The 75 rows will be forecasted with 25 instances in the future(until it reaches 100). This will give as 25 forecasted instances that will be compared to the 25 instances from the actual data.

4. At this step, the result of the forecasted data was sorted in Excel with the rows of revisions, actual result, SVM prediction (SVMpred), ANN prediction (ANNpred), SVM difference (SVMdiff), ANN difference (ANNdiff). The SVM/ANN predictions was the forecasted result. Next, the forecasted result was compared with the actual result (real data) and the difference was obtained as SVMdiff/ ANNdiff (table 5).

5. The result of the 25% forecasted instances was sorted in a table and was compared to the actual instances of data. A calculation of the differences in numbers between the predicted data and the actual was implemented to identify how big the prediction was to the "real" data. A list was created of the difference between the real data and the SVM and ANN prediction data. To find out how accurate the forecast is, a comparison between the real data and the predicted data is conducted, this can indicate how close/ distant the prediction is.

6. A T-test will be conducted to examine if the two approaches (SVM and ANN) has difference in prediction accuracy. The result of the p-value and the t-value will support the decision of the hypothesis rejection and display how big difference it is between the accuracy of the approaches.

### *5.2 Experiment Results and Analysis*

The result of the t-test will display if there exist a difference of accuracy for the approaches SVM and ANN when implemented on the data. This will support a decision to reject or accept the null hypothesis of similar accuracy of SVM and ANN. Also, the outcome of result will highlight if the alternative hypothesis A certain approach, SVM or ANN gives higher accuracy in forecasting is valid.

The result of the p-value for all models was $< 0.05$, this supports the decision to reject the null hypothesis as a significant result.That is to say, the alternative hypothesis is true and shows a significant difference in accuracy between the two approaches. Notably, the comparison of the mean of the two approaches shows that ANN has a lower mean. The data sample estimates SVM mean as 8398 and the ANN as 1875. The comparison of the means is estimated from the difference of predicted LOC to the actual LOC (described in section 5.1).

The result of the t-test above answers the research question 1.2 that is to say, how do the applicable approaches perform regarding prediction accuracy? As the test is significant, the presumption can be made that the mean of the ANN approach is lower than corresponding SVM mean value.

Also, another t-test was conducted on the models based on maximum 7 revisions which includes each file model. The p-value for this test was > 0.05 which is not significant to reject the null hypothesis. However, the sample estimates mean of SVM as 332 and ANN as 1819. Comparing the result of those means to the test above where all models were included, it can be concluded that the mean of SVM had great change from 8398 to 332. The ANN on the other hand did not have any remarkable difference. Also, the mean value of SVM is lower than the ANN mean value, but due to the insignificant result, this cannot be accepted to answer research question 1.2.

From the findings, recurring discoveries were made from analysing the outcome and the source files during the experiment. Several of the files had sudden spikes of both code growth as well as increase in the complexity. The problem here is that the underlying events causing this is not included by the forecasting and therefore expecting these jumps to happen again. However testing to remove some of the outliers where sudden code escalation occurs and then returns to the same value as before, show greater result since the events causing it, meaning if the causing event is not persistent. Aforementioned the quick hikes that appears creates problems for the forecasting throughout all tested files. Taking in to account that these events just happens once or that the future increase will be less, one suggestion can be to start the training data that the forecast is made of to start after these cases to receive a more accurate result.

To get a better understanding and to create better forecasts on must know some of the background of the sudden growths and reductions. The components sometimes are changed by merging code from branches or copying code from other modules and therefore increasing rapidly and therefore a choice of determining if it should be included in the training data once more.

*Revisions and accuracy*

To identify relations and patterns in the data set, three figures will be created. The figures will help to illustrate if there is correlations between accuracy, model size, revisions and amount of predicted revisions.

The following figure 3 is a result from examining the relation between the size of the files i.e. how does the amount of revisions affect the forecasting accuracy. The table is produced by the y-axis indicating the amount of revisions, which will be counted by the revision at the point where the forecasting begins, namely at 75%. The x-axis represents the mean of the predicted LOC values from each file divided with the actual result. Important to know is that the value of 1 in the x-axis represents 100% accuracy which indicates that the means of the forecasting and actual is the same. The calculations divide the actual LOC with the predicted LOC and a percentage is given to show how close or far the prediction was, e.g., if the value of the accuracy exceeds the value of one on the x-axis, the result indicates that the prediction has overestimated the actual LOC value. Assume that the predicted LOC mean is 2000 and the actual LOC mean is 1000, this will give the result two on the x-axis.

To conclude that the accuracy is made from computing the actual mean and dividing it with the forecasted method of SVM and ANN. The reason for the table to exceed 1 i.e. 100% accuracy and not using absolute values is to inspect if the forecast is overestimating or vice versa compared to the actual outcome.

For the files the size and amount of forecasting is based removing 75% and then forecasting the rest 25% as mentioned before. This results in the amount of revisions forecasted are different depending on the file size.

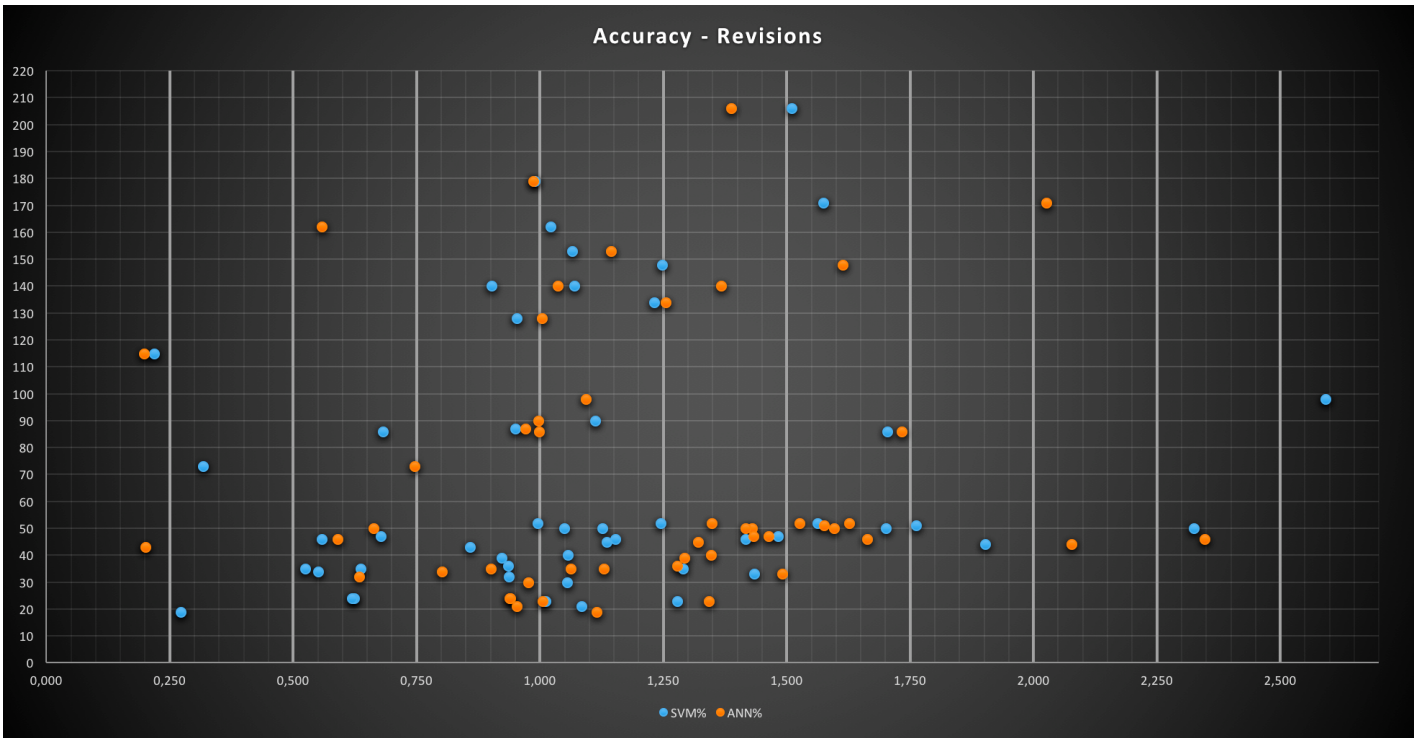



Figure 3. Accuracy for amount of revisions

A closer look at the figure 3, the accuracy level between 0,750 and 1,250, the area close to 1,0 (100% accuracy) gives an indication that SVM performance with higher accuracy, without considering the revision size in the analysis. However, if the revision size is considered (above 100 revision), the SVM is more composed in contrast to ANN which have a bigger spread.

The visualisation shows that there exists a higher spread with files containing revisions below 75 revisions. The accuracy of these files varies from around 50% and until about 175%. In regard to files where revisions were greater than 100 revisions there existed a more concentration of the result between a smaller interval. Still outliers were present in the range mentioned.

The descriptive figure 3 x-axis represents the amount of forecasted instances, in this case revisions. The primary y-axis (left), represent accuracy where 1,00 is 100% accuracy, i.e., 1,500 is 50% accuracy with an overestimation of 50% to the actual LOC prediction. The secondary y-axis (right) presents the amount of models that were available for the forecasting level on the x-axis.

As shown in the figure the descending grey line is another representation of how many models are being used for the forecasted number. The aim of this figure was to analys correlation between amount of models and predicted LOC.

From interpreting the figure 4, an observation can be made that ANN has a more vast distribution when the revision amount is minor i.e. revisions < 15. In larger files the accuracy prediction tend to overestimate the actual value for both implementations. Still, it does not seem to be any clear patterns and most of the predictions are difficult to determine from analysing the figure descriptively.

*Outliers*

Two data points that is excluded in the figure (3) was considered as outliers with very poor prediction accuracy. The two data files had major changes in LOC, i.e., from 3000 LOC to 18000, then after 4-5 revisions going back to 3000 LOC. This is according the analyses of the data files a major change in the LOC pattern and gives the forecasting approach a difficulty to predict.

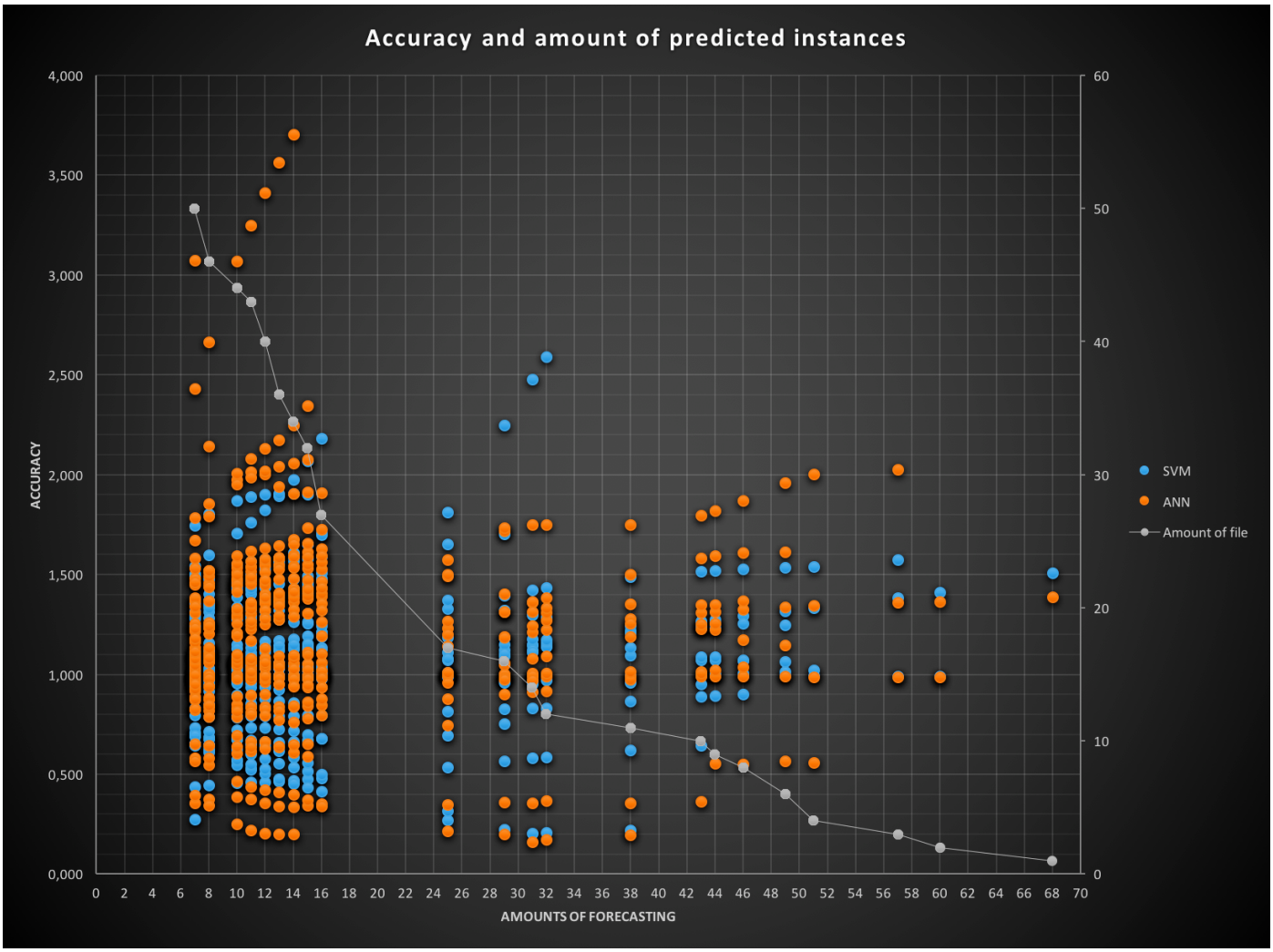
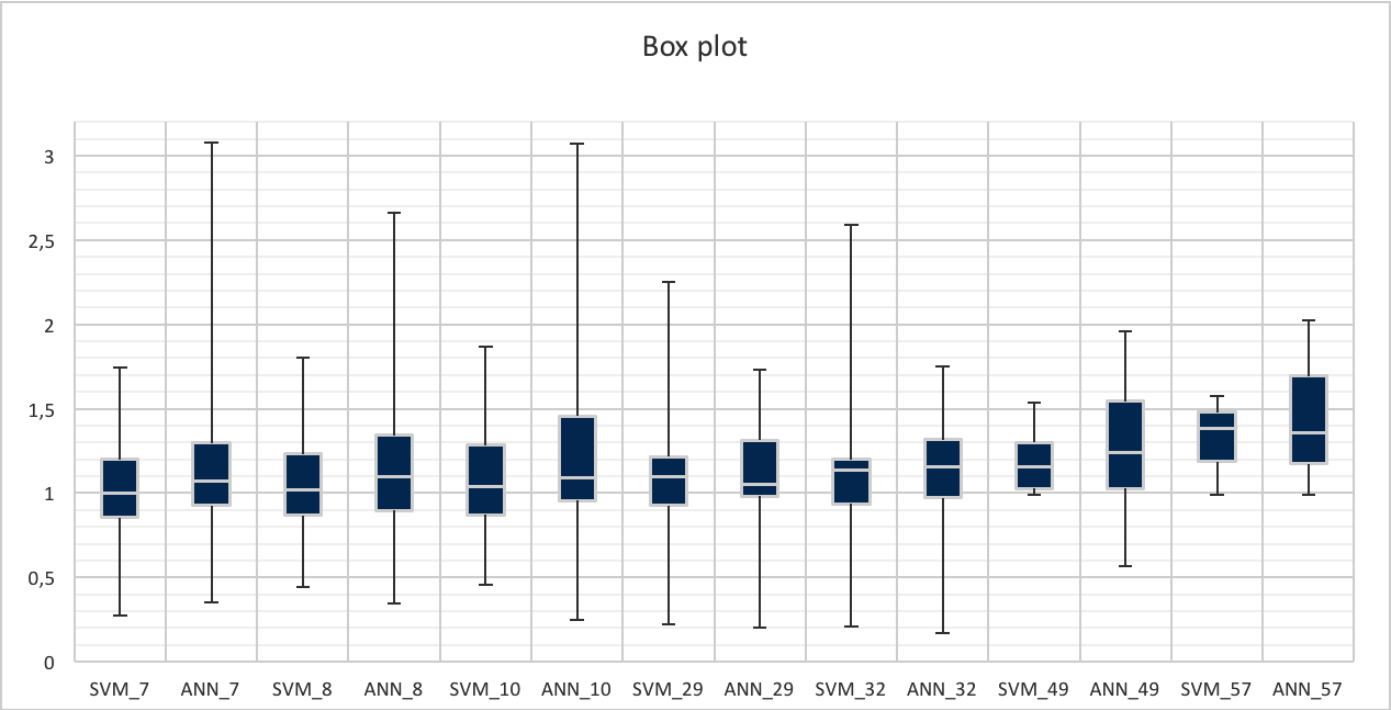



Figure 4. Accuracy for each model in relation to forecasting units




Note: The numbers (7, 8, 10 etc) after the approach abbreviation represent the number of forecasted revisions.

Figure 5. box plot of  for amount of revisions

From the box-plot (figure 5) their can be two main assumptions. Firstly, the wide spread of accuracy is present, even though the levels of forecast are different, the outcome varies a lot. This being said the files with fewer revisions (less forecasts)  i.e. below 30 forecasts still have a more accurate median when compared to the median of the larger files.

## 6. DISCUSSION

In this thesis, the methods have been implemented using the methods that were found mostly used in the current literature and applied to the case study data. From figures 2 and 3, the result shows that there is a vast spread for both SVM and ANN throughout the accuracy level. The forecasting is more often overestimating in comparison to the result but still representation of underestimation is present of both methods.

Looking at figure 2, which illustrates all the possible forecasting that each model is capable of indicates the same result, namely that the wide range of outcomes are depicted for both methods once again.

The rapid changes of code in this context creates problems for the forecasting without knowing to which extent is should rate the escalating swift. By introducing another nominal factor based on the expertise of either the developers or project managers declaring a what type of event that caused the outcome of change. By using the nominal factor both the qualitative element will be taken into consideration but also assisting the forecasting outcome by highlighting the severe events.

Earlier work also indicate that the events triggering the cause of change is an important factor that needs to be used more. Gerlien et al. [33] made an empirical study of SVM and ANN on financial trading and comparing the current findings within this area. From their research, an occurring pattern showed that the majority of studies that were analysed had modest results. Since the arguments were unanimous that just by the usage of numeric values is not enough due to other complex factor that is affecting the result. Even though the usage of machine learning has been adopted and studied more in the context of stock and finance the findings for this study has correlations.

Seeking to improve the accuracy of the forecasting result, a review of the data set can be made to find possible underlying factors causing the poor result. The pure numeric values have been seen with poor results in comparison with a blend of both numeric and nominal values, that has been discussed before in the literature [20]. Using nominal factors that are based on qualitative background in combination with numeric values has shown superior results [20]. Experiences from this study has exposed that files containing code with major changes has produced less accurate results. Therefore additional factors could help the forecasting phase by adding a new aspect when predicting the outcome.

Finally to say, the accuracy level of prediction is depending the information provided to the machine learning algorithm. The data consisting numeric values and nominal values could provide the best prediction outcome, certainly if there is major variations occurring in the data. However, if the data trend does not contain major variations, one type of data (numeric, nominal etc) can be enough to have prediction with high accuracy.

## 7. CONCLUSION

We set out to explore the current machine learning approaches that are mostly implemented in today's research. In order to complete this task we firstly needed a starting point to thoroughly investigate what methods are used and implemented to track which industrial fields are currently using them. From our premise with RQ1 in mind, the support vector machine and artificial neural network were identified as the most commonly used algorithms from the completed literature review. In our context the two aforementioned techniques were decided to be used in an experimental environment where forecasting ability will be assessed.

The data mining tool WEKA was used to perform the experiment, starting with a pre-test for training of the two methods before implementing the test on the rest of data models. Unit of analysis for this experiment was time series data from software measurements, received in an earlier study. The results of the t-test showed that there is a significant difference between the two implemented machine learning approaches. ANN had a lower mean accuracy than SVM.

## 8. FUTURE RECOMMENDATION

For future research, the use of mixed attributes, namely with both numeric and nominal data may contribute to more accurate forecasts. New tests would give additional insight to the applicability of forecasting software development measurements with the use of machine learning. Also attribute selection, i.e., the data that is used is imperative when creating models, and for testing purposes to distinguish which combinations are most efficient.

In this paper the use of size, with the attributes of lines of code and block count as well as a complexity perspective are general for most development processes and can be acquired as a basis. While new artefacts and measurements could result in other outcomes that might be of interest. Another factor could be to use larger data sets to possibly incorporate big data concepts to investigate in patterns that might result in higher accuracy.

## 9. REFERENCES


[1] M. Lesk, "Big Data, Big Brother, Big Money", IEEE Security & Privacy, vol. 11, no. 4, pp. 85-89, 2013.

[2] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, Wei-Yang Lin, Intrusion detection by machine learning: A review, Expert Systems with Applications, Volume 36, Issue 10, December 2009, Pages 11994-12000, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2009.05.029.

[3] Karel Dejaeger, Wouter Verbeke, David Martens, Bart Baesens, Data Mining Techniques for Software Effort Estimation: A Comparative Study, IEEE Transactions on Software Engineering, vol. 38, no. 2, pp. 375-397, March- April, 2012

[4] Cagatay Catal, Software fault prediction: A literature review and current trends, Expert Systems with Applications, Volume 38, Issue 4, April 2011, Pages 4626-4636, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2010.10.024.

[5] Tak-chung Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence, Volume 24, Issue 1, February 2011, Pages 164-181, ISSN 0952-1976, http://dx.doi.org/10.1016/j.engappai.2010.09.007.

[6] Kattan, A., Fatima, S., & Arif, M. (2015). Time-series event-based prediction: An unsupervised learning framework based on genetic programming. Information Sciences, 301, 99–123. http://doi.org/10.1016/j.ins.2014.12.054

[7] Storey, J., Quintas, P., Taylor, P., & Fowle, W. (2002). Flexible employment contracts and their implications for product and process innovation. The International Journal of Human Resource Management, 13(1), 1–18. http://doi.org/10.1080/09585190110092758

[8] Meyfroidt, G., Güiza, F., Ramon, J., & Bruynooghe, M. (2009). Machine learning techniques to examine large patient databases. Best Practice and Research: Clinical Anaesthesiology, 23(1), 127–143. http://doi.org/10.1016/j.bpa.2008.09.003

[9] Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J. (2015). A Survey on Data Mining Techniques Applied to Electricity-Related Time Series Forecasting. Energies (Vol. 8). http://doi.org/10.3390/en81112361

[10] Bose, I., & Mahapatra, R. K. (2001). Business data mining — a machine learning perspective. Information & Management, 39(3), 211–225. http://doi.org/10.1016/S0378-7206(01)00091-X

[11] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: Literature review and challenges. International Journal of Distributed Sensor Networks, 2015(i). http://doi.org/10.1155/2015/431047

[12] Malhotra, R. (2015). A systematic review of machine learning techniques for software fault prediction. Applied Soft Computing Journal, 27, 504–518. http://doi.org/10.1016/j.asoc.2014.11.023

[13] Carpinteiro, O. A. S., Leite, J. P. R. R., Pinheiro, C. A. M., & Lima, I. (2012). Forecasting models for prediction in time series. Artificial Intelligence Review, 38(2). http://doi.org/10.1007/s10462-011-9275-1

[14] Ramirez-Amaro, K., & Chimal-Eguia, J. C. (2007). Machine learning tools to time series forecasting. 2007 Sixth Mexican International Conference on Artificial Intelligence, Special Session, MICAI. http://doi.org/10.1109/MICAI.2007.42

[15] Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(TR/SE-0401), 28. http://doi.org/10.1.1.122.3308

[16] Hilbert, M. and Lopez, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025), pp.

60-65.

[17] M. Unterkalmsteiner, T. Gorschek, A. Islam, Chow Kian Cheng, R. Permadi and R. Feldt, "Evaluation and Measurement of Software Process Improvement—A Systematic Literature Review",IEEE Transactions on Software Engineering, vol. 38, no. 2, pp. 398-424, 2012.

[18] Stol, K.-J., & Fitzgerald, B. (2015). A Holistic Overview of Software Engineering Research Strategies. Proceedings of the 3rd International Workshop on Conducting Empirical Studies in Industry, 8. http://doi.org/10.1109/CESI.2015.15

[19] Xiaozhe Wang, Smith-Miles, K., & Hyndman, R. (2009). Rule induction for forecasting method selection: meta-learning the characteristics of univariate time series. Neurocomputing, 72(10-12). http://doi.org/10.1016/j.neucom.2008.10.017

[20] Yoo, P. D., Kim, M. H., & Jan, T. (2007). Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation. International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), 2, 835–841. http://doi.org/10.1109/CIMCA.2005.1631572

[21] G. Box and G. Jenkins, Time series analysis. San Francisco: Holden-Day, 1976.

[22] Garcia, M. P., & Kirschen, D. S. (n.d.). Forecasting system imbalance volumes in competitive electricity markets. In IEEE PES Power Systems Conference and Exposition, 2004. (pp. 1115–1122). IEEE. http://doi.org/10.1109/PSCE.2004.1397617

[23] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," Comput. Educ., vol. 53, no. 3, pp. 950–965, Nov. 2009.

[24] Vanajakshi, L., & Rilett, L. R. (2007). Support vector machine technique for the short term prediction of travel time. In IEEE Intelligent Vehicles Symposium, Proceedings (pp. 600–605). Retrieved from http://

www.scopus.com/inward/record.url?eid=2-s2.0-47849099540&partnerID=tZOtx3y1

[25] Gerdes, M. (2013). Decision trees and genetic algorithms for condition monitoring forecasting of aircraft air conditioning. Expert Systems with Applications, 40(12), 5021–5026. http://doi.org/10.1016/j.eswa.2013.03.025

[26] D. Lindsay and S. Cox, “Effective probability forecasting for time series data using standard machine learning techniques,” Pattern Recognit. Data Min., vol. 3686, pp. 35–44, 2005.

[27] Radzuan, N. F. M., Othman, Z., & Bakar, A. A. (2013). Uncertain Time Series in Weather Prediction. Procedia Technology, 11(Iceei), 557–564. http://doi.org/10.1016/j.protcy.2013.12.228

[28] Choi, T. M., Hui, C. L., & Yu, Y. (2011). Intelligent time series fast forecasting for fashion sales: A research agenda. Proceedings - International Conference on Machine Learning and Cybernetics, 3, 1010–1014. http://doi.org/10.1109/ICMLC.2011.6016870

[29] Sapankevych, N., & Sankar, R. (2009). Time Series Prediction Using Support Vector Machines: A Survey. IEEE Computational Intelligence Magazine, 4(2), 24–38. http://doi.org/10.1109/MCI.2009.932254

[30] S. F. Crone, S. Lessmann, and S. Pietsch, “Forecasting with Computational Intelligence - An Evaluation of Support Vector Regression and Artificial Neural Networks for Time Series Prediction,” in The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006, pp. 3159–3166.

[31] V. Basili, R. Selby and D. Hutchens, "Experimentation in software engineering", IEEE Transactions on Software Engineering, vol. -12, no. 7, pp. 733-743, 1986.

[32] J.Schroeder, C. Berger, M. Staron, T. Herpel & A. Knauss, "Unveiling Anomalies and Their Impact on Software Quality in Model-Based Automotive Software Revisions with Software Metrics and Domain Experts," in The International Symposium on Software Testing and Analysis, Saarland University, Saarbrücken, Germany, July 18–20, 2016, Forthcoming.

[33] Gerlein, E. A., McGinnity, M., Belatreche, A., & Coleman, S. (2016). Evaluating machine learning classification for financial trading: An empirical approach. Expert Systems with Applications, 54, 193–207. http://doi.org/10.1016/j.eswa.2016.01.018