

# MSE PROJECT TITLES

Semester 1, Session 2020/21

**Lecturer: Dr. Chiam Yin Kia**

**Email: yinkia@um.edu.my**

**Title of project:** A predictive model for hand, foot, and mouth disease in Malaysia

**Research Area: Data Mining**

**Objective(s) of project:**

1. To identify the correlation between risk factors and climatic factors that influence the outbreak of hand, foot and mouth disease (HFMD) in Malaysia.
2. To propose a model to predict the outbreak of HFMD in Malaysia based on machine learning algorithms, risk and climatic factors.
3. To evaluate the performance of proposed predictive model using HFMD datasets in Malaysia.
4. To develop a visualization tool to analyse the HFMD datasets.

**Brief description:**

Hand, foot and mouth disease (HFMD) causes significant morbidity in Malaysia. For instance, Malaysia recorded 51,147 cases of HFMD between January 1 and August 14, 2018. Occasionally, it can also lead to brain infection and deaths. Previous studies have indicated an association between the incidence of HFMD with climatic factors. However, the seasonal influence of these factors on HFMD is not yet understood, and the temporal-geospatial patterns are not well documented. The risk factors and climatic factors which may drive a HFMD outbreak remain unclear. This research aims to analyse HFMD cases from 2009 to 2016 in Malaysia to explore the risk factors and seasonal influence of climatic factors on incidence of HFMD by using a predictive model. A predictive model identifies features (or factors) and machine learning algorithms that are relevant for HFMD prediction. The predictive model is of importance in this context of study because it would help to improve healthcare and monitor the outbreak of HFMD in Malaysia. Prediction of HFMD will enable the planning and focusing of resources for control measures. A visualization tool need to be developed to show/highlights all the variables and analytics in the HFMD datasets.

**Expected Outcome:** In this research, a predictive model will be proposed to predict the outbreak of HFMD. Additionally, a tool need to be developed to visualize variables and analytics in the HFMD datasets.

**Tools/Programming languages to be used:** Python, JavaScript, Java, R or any suitable programming languages that can be used to analyse and visualize the HFMD datasets.

**Title of project:** Process reference model/framework to develop machine learning (ML) based systems

**Research Area: Software process improvement**

**Objective(s) of project:**

1. To identify the differences between machine-learning (ML) based systems development and generic software development.
2. To analyse the best practices or steps used to ensure the statistical validity of the ML-components in the ML based systems.
3. To develop a process reference model/framework to integrate ML component development process into generic software development process to ensure the statistical validity of the ML components.

**Brief description:**

Machine learning (ML) will enable cognitive systems to assist us in making good decisions by bringing the right recommendation to us in a more natural and personalized way. Furthermore, many ML algorithms have been applied to analyse big data quickly and automatically in recent software systems development. However, many experienced software engineers are experts in generic software development but they are a novice at ML-based system development and unfamiliar with the methodology of implementing ML components. On the other hand, a novel ML method is typically proposed by a ML researcher or team of researchers. A problem is that the code is written by researchers that may or may not be trained in the discipline of software development to ensure the statistical validity of the ML components.

There are various risks that can be identified and need to be addressed when we integrate ML-based component(s) into a software system. Sculley et al. (2014) have identified several ML specific risk factors that need to be avoided or refactored during ML based system development. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, changes in the external world, and a variety of system-level anti-patterns (Sculley et al, 2014). There is a change of software engineering role in the ML based system development. The development teams need to ensure that the requirements that we expect from ML-based systems can still be met when the ML-based components keep involving. The outputs produced by ML components should always meet the statistical validity specified in the requirements. For example, ML-based system may deal with big business/personal data which is different from dealing with scientific data. There are unchanging scientific laws underlying scientific data while the learned laws using big business/personal data are not only more fluid but also the produced predictive results that may "change" the end users and the underlying learned rules over time.

This research aims to address the following research questions:

- What are the software development practices that could make a big difference when experimenting and testing ML algorithms?

- What are the good ML development practices that are applied by ML researchers to ensure the statistical validity of ML components?
- How to integrate the practices or steps used in ML components development into generic software development processes?

This research aims to propose a conceptual process reference model/framework for ML based systems, that support software engineers, ML researchers, statisticians and/or data analyst in developing and maintaining a ML based systems to ensure the statistical validity of the ML components. The model/framework should consist of a holistic life cycle model for ML based systems, product quality criteria and metrics for ML components, best practices for different stakeholders, and recommendations for action, as well as tools that support stakeholders in developing and maintaining ML-based systems.

**Expected Outcome:**

In this research, a conceptual process reference model/framework will be proposed for supporting ML based systems development to manage the ML component development and ensure the statistical validity of the ML components.

**Tools/Programming languages to be used:**

Any suitable tool that can be used to demonstrate the new process framework.

**Title of project:** Applying text mining to analyse the quality of open source software

**Research Area:** Software quality

**Objective(s) of project:**

- 1. To identify text mining techniques in analysing text content of defect reports.
- 2. To identify significant defect attributes from existing defect reports for assessing the quality of open source software
- 3. To propose a method for analysing the quality aspect of open source software
- 4. To evaluate the performance of the proposed method using the selected significant defect attributes.

**Brief description:**

Open source software (OSS) such as Eclipse, Apache, Mozilla and Firefox have been widely used by individuals and organisations for decades. OSS is usually free according to its license and anyone can contribute to change and improve the OSS code. Although OSS has gone through user testing and peer reviews of the product community, it is still challenging for the OSS developers to claim the overall quality of OSS products. There are still problems that affect the quality of OSS. It is crucial to understand how we can analyse the quality of OSS. Besides performing static code analysis, defect reports reveal the quality information and issues in an OSS project. Defect reports can be analysed using text mining techniques to identify the types of quality problems and types of fix to resolve the quality defects in the OSS. Figure 1 shows the list of software quality attributes defined in ISO/IEC25010 Software Product Quality Model. One quality attribute will be selected from the software quality model (e.g. usability) to be the research focus in this study.

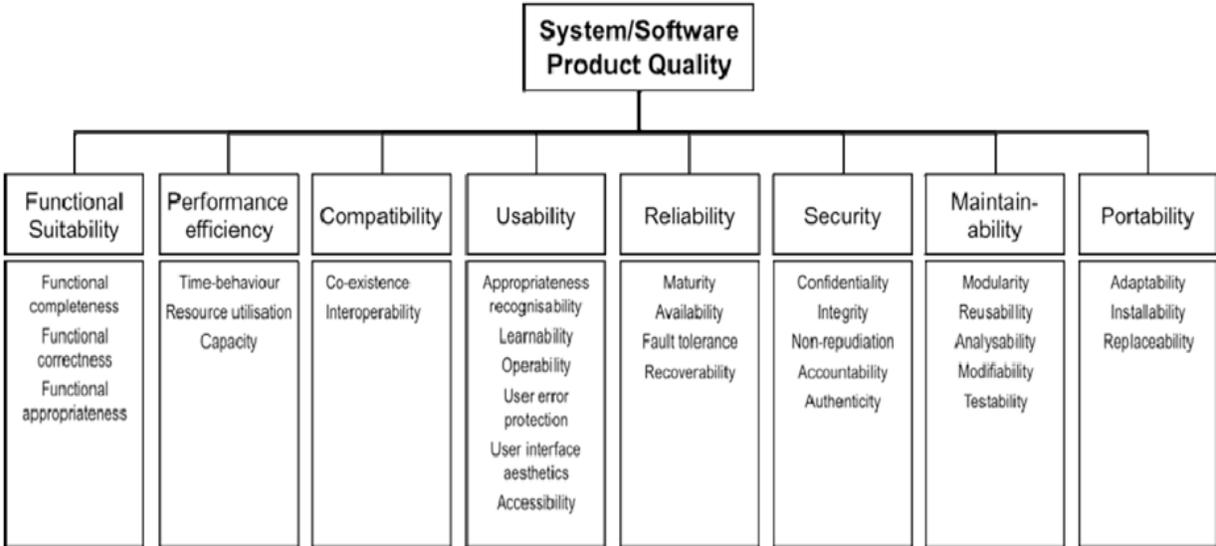


Figure 1: Software Product Quality Model defined in ISO/IEC25010

**Expected Outcome:**

In this research, a method based on text mining is proposed to process the text content of defect reports and analyse the quality aspect of open source software. The scope can be narrowed down to focus on one quality attribute only.

**Tools/Programming languages to be used:**

Any suitable tool that can be used to demonstrate the proposed method

# MSE PROJECT TITLES

**Semester 1, Session 2020/21**

**Lecturer:** Dr. Su Moon Ting

**Email:** [smting@um.edu.my](mailto:smting@um.edu.my)

**Title of project:** Embedding architectural design intent in source code

**Research Area:** Software Design

**Objective(s) of project:**

1. To identify architectural design intent that should be made visible in source code and existing ways of embedding architectural design intent in source code.
2. To derive a solution (either a method/technique/approach) for embedding architectural design intent in source code.
3. To implement the proposed solution into an Integrated Development Environment (IDE).
4. To evaluate the usefulness of the proposed solution in a case study involving software architects.

**Brief description:**

Design intent is usually not visible in the resulting source code. For example, architecture models are generally not visible in the respective code. This creates traceability problem between the two types of artefacts that negatively affects the understandability of code. Following that, this project aims to derive a solution (either a method/technique/approach/etc) for embedding architectural design intent in source code, to implement the solution into an IDE and to evaluate the solution in a case study. Prior to that, this project requires the candidate to conduct a comprehensive review of what types of architectural design intent should be made visible in source code, and existing ways of embedding architectural design intent in source code.

**Expected Outcome:**

1. A comprehensive review of what types of architectural design intent should be made visible in source code and existing ways of embedding architectural design intent in source code.
2. A solution (either a method/technique/approach/etc) for embedding architectural design intent in source code.
3. Implementation of the proposed solution into an IDE.
4. Evaluation results of the usefulness of the solution by software architects.

**Tools/Programming languages to be used:** To be determined. This might involve customising or extending or integrating suitable existing tools.

# **MSE PROJECT TITLES**

**Semester 1, Session 2020/2021**

**Lecturer: Dr. Mumtaz Begum Mustafa**

**Email: mumtaz@um.edu.my**

1. **Title of project:** An Algorithm for Correcting Errors produced by speech recognition system in mobile applications.

**Research Area: Human Computer Interaction, Speech Recognition**

**Objective(s) of project:**

1. To analyse the types of errors commonly occurred in mobile applications' speech recognition system in converting speech to text/command
2. To design and develop an error correction interface for a mobile application's speech recognition system using a suitable algorithm which suggests ways to reduce the user's effort during the human-machine interaction process.
3. To evaluate the performance of the algorithm in correcting the errors produced by the speech recognition system.

**Brief description:**

In recent years, voice based input interface has become popular in smart phone applications. One key component of a Voice user interfaces (VUI) is automated speech recognition (ASR) that enables users' speech to be translated into text. However, the recognition errors of this form of interface are still occurring and are unavoidable due to factors such as the inability of the system to determine the end of speech, or using talking too much and too long.

When a user speaks with the VUI, the speech recognition system generates more than one response to what was said. It then assigns a confidence value and usually picks the one that has the highest confidence value. In simple terms, a confidence value is a percentage that indicates how confident the system is about a particular result. However, a system with high confidence value can still generate wrong result.

To compensate the error, users are required to verify the ASR output and correct the errors. Therefore, simpler and more efficient error correction interfaces have been strongly demanded. The aim of this study is to design simpler user interfaces, develop efficient error correction

algorithms, and suggests ways to reduce user's effort during the human-machine interaction process.

**Expected Outcome:**

Design and the development of the error correction interface using suitable algorithm.

**Tools/Programming languages to be used:**

Java, Python, C++, Speech Recognition toolkit/engine (HTK or Kaldi)

**2. Title of project:** Automatic Speech Recognition System for Bilingual Code-Switch Conversational Speech

**Research Area:** Speech Recognition

**Objective(s) of project:**

1. To investigate the performance of the state-of-the-art bilingual back-end framework used for recognizing code-switching speech.
2. To propose a suitable back-end framework that can cope with the language switches.
3. To develop an Automatic Speech Recognition System of Bahasa Malaysia and English languages using the proposed back-end framework that can cope with language switches.
4. To evaluate and compare the performance of the monolingual and bilingual automatic speech recognition systems.

**Brief description:**

Multilingual speakers switch between languages displaying inter sentential, intra-sentential, and congruent lexicalization based transitions. While monolingual ASR systems may be capable of recognizing a few words from a foreign language, they are usually not robust enough to handle these varied styles of code-switching. There is also a lack of large code-switched speech corpora capturing all these styles making it difficult to build code-switched speech recognition systems. One prominent mechanism induced in the interacting languages is code-switching (CS) which is defined as the continuous alteration between two languages in a single conversation.

CS is mostly noticeable in some minority languages influenced by the majority language or majority languages that have been influenced by globally influential languages such as English and Mandarin, French, etc. Despite the well-established research line in linguistics, robustness of speech-to-text systems against CS and other kinds of language switches have recently received some interest resulting in some robust acoustic modeling and language modeling approaches for CS speech.

This research will propose a suitable bilingual back-end framework and develop a bilingual ASR system which is designed to recognize both Bahasa Malaysia and English. By investigating different bilingual framework, this research will aim to get more insight into developing a more versatile acoustic modeling scheme coping with the language switches.

**Expected Outcome:**

A bilingual back-end framework that can cope with language switches, code switch language model and bilingual acoustic model of Bahasa Malaysia and English.

**Tools/Programming languages to be used:**

Python, Speech Recognition toolkit (HTK or Kaldi)

# MSE PROJECT TITLES

**Semester 1, Session 2020/21**

**Lecturer:** Hasan Kahtan

**Email:** hasankahtan@um.edu.my

**Title of project:** Impact Analyze of Software Security Attributes using TOPSIS Method

**Research Area:** Software Engineering / Software Testing / Software Security

## **Objective(s) of project:**

- To analyze the impact level of software security attributes pertaining to software development.
- To develop a guideline for secure software development.
- To evaluate the trustworthiness of the proposed guideline's using TOPSIS method.

**Brief description:** Developing a secure software became the core aim of software companies. To achieve this aim developer should embed the software security attributes along with their software coding. Software security attributes namely, availability, reliability, confidentiality, integrity, safety and maintainability. This research intent to analyze the importance of software security attributes pertaining to development phase. The analyze will be carried out through reviewing current studies related to software security attributes and the tools used to measure them such as JMeter, and OpenVAS. Then come out with a trustworthy guideline to address these security attributes into software development. Finally, evaluate the trustworthy guideline using as Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method. TOPSIS is a multi-criteria decision analysis method.

**Expected Outcome:** A trustworthy guideline for secure software development.

**Tools/Programming languages to be used:** Java and other suitable languages.

# MSE PROJECT TITLES

**Semester 1, Session 2020/21**

**Lecturer: Raja Jamilah Raja Yuso**

**Email: [rjry@um.edu.my](mailto:rjry@um.edu.my)**

**Title of project:** Designing a model for new approaches of user interface Interaction paradigm

**Research Area: Human Computer Interaction**

**Objective(s) of project:**

- To analyze and identify interaction paradigm in relation to tasks to be accomplished through user interfaces
- To designing a model for new interaction approaches in accomplishing user interface interaction tasks
- To develop and evaluate the model above

**Brief description:**

The keyboard, the mouse and touch pads are existing interaction medium existed since the 1960s. With the advancement of technology, the user interface interaction paradigm still relies heavily on keyboard and mouse as well as touch screen in mobile devices in its original medium and design only differing in the electronic components used for the medium. Introducing new approaches in accomplishing normal tasks in a user interface can be done using the existing interaction paradigm that can increase effectiveness, efficiency and user satisfaction. For example instead of using the mouse click event for clicking on a particular data that we see, mouse click event can be used to control far away computer display through bluetooth or infrared signal to do task such as transferring data from one computer to another, flipping pages in reading an ebook, closing or shutting down a computer / device.

Candidate needs to analyze and identify interaction paradigms to accomplish certain tasks using Goals, Operators, Methods and Selection (GOMS) model. A new interaction model is to be designed based on the GOMS analysis using the Cogulator tool (Savage-Knepshild, 2014). Next students need to develop and evaluate the model above. The development of the interaction model can use existing mouse and keyboard preferably wireless ones and/or other devices such as sensors, arduino nano, raspberry pi zero, others. The evaluation method should focus on the usability of the system in terms of effectiveness, efficiency and user satisfaction.

**Expected Outcome: A prototype of the** new approach in accomplishing tasks using existing interaction paradigm

**Tools/Programming languages to be used:** Web-based platform, Eclipse /Java /Python

**Reference:**

Savage-Knepshield, (2014). Human Factors and Ergonomics Society: HFES Bulletin, Vol 57, No. 6, June 2014. Accessed from:

[http://cogulator.io/resources/Cogulator\\_Interview\\_HFES\\_Bulletin.pdf](http://cogulator.io/resources/Cogulator_Interview_HFES_Bulletin.pdf)

<http://cogulator.io/primer.html>

**Title of project:** Analyzing Maturity Level of Agile Machine Learning Projects

**Research Area:** Software Process

**Objective(s) of project:**

- To analyze Maturity Level of current practices of Agile Machine Learning Projects
- To identify machine learning projects to be analyzed in terms of maturity level of current practices in Agile Machine Learning Projects
- To design an agile machine learning maturity level model based on currently existing model
- To apply and develop tools for analyzing the above model

**Brief description:**

Maturity models can guide organizations in providing the directions of the organization's working strategy. A maturity model is a conceptual framework that comprises a collection of best practices that help organizations to improve their processes in a particular area of interest. On the other hand, machine learning projects are known to be highly iterative such as shown in Figure 1 (Jordan, 2018) and therefore could use agile development practices to help speed up the project. This research is to analyze the maturity level of current practices in agile machine learning projects. It is suggested in developing the tools, the candidates should use existing agile maturity models as a basis for agile practices and extend it with practices that are key to machine learning projects (Oktay and et. al, 2016). The model should be developed and refined with domain experts using the Delphi or any other suitable techniques. The model is to be developed and refined with domain experts using the Delphi technique.

Case studies should be used to evaluate the model to assess the level of Maturity Level in the Agile Machine Learning projects.

**Machine Learning Development Lifecycle**

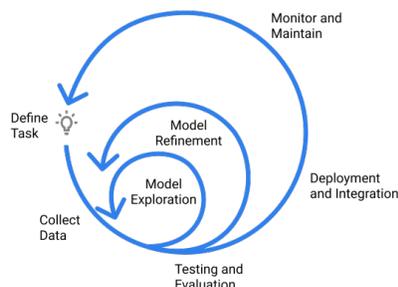


Figure 1: Example of Machine Learning development Lifecycle (Jordan, 2018)

**Expected Outcome: Analysis of Maturity Level, Full or semi automated or/and formulated tool**

**Tools/Programming languages to be used:** Maturity Level tools, (Proença & Borbinha, 2018), Eclipse /Java /Python

**References:**

Jordan Jeremy (2018), Organizing machine learning projects: project management guidelines.  
<https://www.jeremyjordan.me/ml-projects-guide/>

Proença, Diogo & Borbinha, José. (2018). Maturity Model Architect: A Tool for Maturity Assessment Support. 42-51. 10.1109/CBI.2018.10045.

[Oktay Turetken](#), [Igor Stojanov](#), [Jos J. M. Trienekens](#) (2016), Assessing the adoption level of scaled agile development: a maturity model for Scaled Agile Framework, Journal of Software: Evolution and Process, Volume 29, Issue 6 2016.

<https://onlinelibrary.wiley.com/doi/10.1002/smr.1796>

# MSE PROJECT TITLES

Semester 1, Session 2020/21

Lecturer: Dr. Adeleh Asemi Zavareh

Email: [adeleh@um.edu.my](mailto:adeleh@um.edu.my)

## **Title of project: Knowledge Management Automation in Knowledge Based Inference Systems**

Status: New / Approved (please strike through where applicable) New

Research Area: Expert systems

Objective(s) of project:

Main Objective: to automate the process of designing membership function in inference systems

Objective 1: to determine knowledge representation features

Objective 2: to determine relation between features and linguistic variables

Objective 3: to provide a recommender system for representing knowledge based on linguistic variables

Objective 4: to apply proposed recommender system in Matlab fuzzy tool

Brief description: This project proposes a recommender system for generating fuzzy membership functions of inference systems under different platforms especially, fuzzy inference systems and controllers which are implemented through Matlab. For implementation of expert system knowledge engineer should design the membership function and there is not any rule or references that help him/her to decide regarding membership functions. Most of knowledge engineers select defaults of fuzzy tools for designing membership function for each input or output factor. However, accuracy of system depends on designing proper membership function. In this project we will provide an automatic system for generation of proper membership functions based on linguistic variables. Proposed system improves the usability and accuracy of current expert system tools.

Expected Outcome: In this project we propose a recommender system for generation of membership functions.

Tools/Programming languages to be used: Java

**Title of project: Implementation of a Platform for Occupational Therapy in Children with Sensor Processing Disorder (SPD)**

Status: New / Approved (please strike through where applicable) New

Research Area: Human Computer Interaction

Objective(s) of project:

Main Objective: to provide a platform for detection and treatment of sensor processing disorder

Objective 1: to determine SPD detection factors.

Objective 2: to classify different types of SPD.

Objective 3: to determine the best treatment for each type of SPD.

Objective 4: to provide a platform for detection of SPD type and recommendation of best sensory diet.

Objective 5: to evaluate the proposed platform.

Brief description: This is an interdisciplinary research that needs a comprehensive data collection for detection and treatment of SPD and ASD. The end users are therapists, parents and children with SPD or ASD. Since SPD is differ from one person to another, this project provide a personal treatment based on situation of each SPD patient. This project provides computer application and software as recommender systems for determination of proper sensory diet or exercise based on children disorder type and severity. This project helps occupational therapists for treatment of children with Sensor Processing Disorder (SPD) and Autism Spectrum Disorder (ASD).

Expected Outcome: Application/Software for occupational therapy of children with SPD. The existing mobile applications of lose weight diet are examples of expected outcome for this project.

Tools/Programming languages to be used: Java or any other programming language.