Computer Science @ RIT
B.Thomas Golisano College of Computing & Information Sciences

Rochester Institute of Technology,
Department of Computer Science

CSCI-630 FOUNDATIONS OF INTELLIGENT SYSTEMS

Research and Development project

Research, design, and investigation of AI and machine learning applications

Date due: see a separate deadline for each part on MyCourses

Project learning outcomes. As a result of this project you should be able to:

- Demonstrate an understanding of the key principles involved in artificial intelligence (AI) application development.
- Demonstrate knowledge of expert systems and their design and implementation.
- Demonstrate knowledge of neural networks and machine learning techniques and their design and implementation
- Demonstrate skills in conducting research in artificial intelligence and computer security domains
- Demonstrate an ability to design and implement a software engineering project in order to conduct the research and development activities stated above.

This semester you will run ONE project, which consists of a few parts. You will have to submit all the parts separately. Those parts will be assessed separately also. However, unless and until you finish one part, you will not be able to start the next one, so PLEASE keep your work on schedule.

You can do your projects either individually or in small groups of two or three students. The project has been designed for small groups and might require substantial efforts. The requirements for individual and group projects may differ slightly but generally will be very similar. However, I am expecting a higher quality submission from a group of three in comparison to an individual. Also, the penalty for late submission/resubmission will be different: for individual projects, it will be 10% grade reduction per day and for group projects, it will be 20% grade reduction per day. If you do your project in a group, please, coordinate your efforts and do it on time. I suggest you to have an initial discussion about your work assignments and schedule your work to finish at least a few days before the deadline. However, I am leaving all organizational aspects to your decision.

The project requires learning the problem, designing and implementing the **security evaluation system** based on the expert system and machine learning technology you are studying in this class.
The general content of the whole project will be as follows:
Part 1: Bibliography and patent research, decision making and project specification development
Part 2: Expert system design and implementation
Part 3: Machine learning app design and implementation
Please, note that this is *a research project*. I do not want you to follow up my instructions only. I will really appreciate your ideas and their implementation in conducting this project and delivery of its results.
I am attaching the paper "Artificial Intelligence Based Design and Implementation of Data Quality and

Security Calculus on Android Mobile Devices" (see attachment 2) that describes our work that has been done and the process you have to follow up. You can use this paper but as it has not been published yes, please, keep it **confidential**. You are NOT allowed to distribute any part of this paper and use its content outside of this class until it is published.
Specific content and submission requirements for each part are provided in the later sections.
Also, please, see the project grading scheme in attachment 1.

You have to design an application to be used for evaluating the security of the specific computer environment. The paper mentioned above describes an evaluation of the Android smartphone. You are welcome to design an app that performs a similar evaluation. If you produce a working Android app, you may get the bonus (see more information below). However, you can develop an application that evaluates a computer security of another platform (e.g. your Windows desktop, or your Mac laptop) or one that could be used for security evaluation of a class of systems. The required knowledge could be acquired from the papers attached, papers reviewed (see project 1 description) and various documents including Security Metrics Guide for Information Technology Systems, produced by the Computer Security Resource Center of the National Institute of Standards and Technology (NIST). You can download this document from http://csrc.nist.gov/publications/nistpubs/800-55-Rev1/SP800-55-rev1.pdf
This document provides guidance on how an organization, through the use of metrics, identifies the adequacy of in-place security controls, policies, and procedures. It provides an approach to help management decide where to invest in additional security protection resources or identify and evaluate nonproductive controls. It explains the metric development and implementation process and how it can also be used to adequately justify security control investments. The results of an effective metric program can provide useful data for directing the allocation of information security resources and should simplify the preparation of performance-related reports. However, you have to understand that those metrics are mainly applicable for large computer systems. For a smartphone environment, another set of metrics might be needed (see the attached paper for the examples).
Your application may automatically read some system parameters (preferably) or request the users to input some data characterizing their computer systems and as an output it should produce some summary evaluation of the security in some scale (e.g. "poor – excellent" or 0-10) and give the user some advice what exactly needs improvement. However, again, in producing security evaluation app for a smartphone, those metrics are supposed to be collected automatically.
Please, note that you have to produce a hierarchical data quality evaluation system. This requirement means that you have to use metrics from at least two domains.

Course Project Part 1.
Big Data Quality Evaluation Research

Due: see MyCourses

The main delivery here will be your research paper, 2-3 pages long for individuals, 6-7 for groups of three students. Please, note that I am assuming normal formatting with about 1" margins and the font size of 11 points in IEEE template available on the department website.

The research topic: Hierarchical expert system for security evaluation and its implementation on an Android smartphone

As you see there are three major topics in your research:
1) Computer security evaluation,
2) Design of the hierarchical expert system for this application,
3) Expert system implementation on Android mobile devices.

To help you in your research kick off, I am attaching the paper "Artificial Intelligence Based Design and Implementation of Data Quality and Security Calculus on Android Mobile Devices**"** (see attachment 2) that describes our work that has been done and the process you have to follow up. You can use this paper but as it has not been published yet, please, keep it confidential. You are NOT allowed to distribute any part of this paper and use its content outside of this class until it is published.

Your research paper could be based on research literature search, analysis, and synthesis of some conclusion. Here you have to find at least 2 papers for one person project or 4 papers for 3 person project, to investigate and study them, analyze them and produce your conclusions. Please, note that I do not want you to re-tell the content of the papers. If you do this, please, keep it really short. I want you to provide a comparative analysis of the methods and results, which you learned from the publications and your evaluation of the results' potential. Also, as you see the attached paper contains the references to a number of publications. You should use none of those. You have to find and analyze those that are not included in the list of paper's citations. The main attention should be paid to the analysis of the research results in particular publications.

In a conclusion of your paper, you have to specify the content of your security evaluation system or app, which metrics and groups you are going to use and the ways how you are going to produce the rules of your expert system and implement them.

THE MAIN RESULT: In the conclusion of your project 1 you have to come up with the list of metrics you are going to use in your evaluation system and the ways how you are going to get data necessary to measure and merge them together in part 2 and the machine learning technique (and its brief description) of security evaluation app implementation in part 3.

Practically, I want you to produce the project specification for your work in project parts 2 and 3 (about 1 page long). You have to research and identify the tools you are going to employ in your design. For example, in project 2 you can choose between CLIPS and Py-CLIPS. In project 3, you can use design packages such as Weka or Matlab as well as ML libraries such as TensorFlow, Caffe, or PyTorch.

Also, please, see the project grading scheme and note that the grader will really appreciate your novel and interesting ideas on security evaluation system. **Please, specify your novel ideas in a separate section**.

Submission Requirements for Project 1.

(1) Submission deadline is 11.59 pm on the date specified on MyCourses
(2) Late submission or resubmissions will be penalized. Note that a delay longer than one hour after midnight will be counted as a day.
(3) Please, submit your project by uploading your files into MyCourses folder Project 1.

You have to submit:

a file named pr1_yourname(s).* should contain your paper in Word or pdf formats only.

DO NOT submit an archive file here, it will not get accepted. Please, note that your submission will be automatically checked for plagiarism.

Course Project Part 2.

Building an expert system

Due: see MyCourses

General Contents

1) You have to design, develop (write a code), run, test and evaluate the software for an expert system, which could be applied for solving a specified set of application problems or a specific problem (you had to produce problem and project specification in project 1)
2) You have to try to improve your product's performance (speed and memory consumption) and make it more reliable and more secure. You have to develop and submit a proper documentation (see the project report section).
3) You have to document :
   a) the product itself and guidelines how to use it,
   b) testing, you performed, and how you evaluated the product based on test results,
   c) the development process (for group projects only): how you distributed the workload, who did what, how much time you spent on different aspects, etc.
4) Developing of a nice GUI is not directly required for one-person project but it could be a plus and you might get some bonus points for really outstanding solutions

Project work:

Step1: **Identify the problem and analyze the knowledge to be included into the system**
Just about any field of activity involves expertise of some kind – it could be described with a more or less formal model or as a rule-of-thumb expertise. Therefore, you will have an opportunity to employ an expert system practically in every domain and improve an efficiency of the decision making process. However, you should possess or be able to gather a sufficient amount of knowledge in the specified domain. Also, keep in mind that in some areas an expert system application could produce more benefits than in others.
You may copy your specification from part 1 or modify it based on your further research and the feedback received.

Step 2: **Choose an implementation tool**
Depending on the type of an expert system you are designing, different tools could be employed. With the rules based expert system, you will need a system shell to fill in with a specific knowledge.
You have three options:
Option 1: Use JESS, the rule engine for Java platform, developed by Ernest Friedman-Hill at Sandia National Laboratories in Livermore, CA and available for free download at http://herzberg.ca.sandia.gov/jess/. Keep in mind that a free download version is valid for one month only. Also, this website lately has become unreliable and may come up randomly. If you decide to use this product, give yourself plenty of time to get it.
Option 2: Use CLIPS, the rules library for C language implementation available for download from http://www.clipsrules.net . Some adds-on for other languages are available too.
Option 3: Use PyCLIPS, an extension module for the Python language that embeds full CLIPS functionality in Python applications. This means that you can provide Python with a strong,

reliable, widely used and well documented inference engine. You can download it from http://pyclips.sourceforge.net/web/
Option 4: Design and develop the tool (shell) by yourself. This will include writing a code in any reasonable programming language but not Java or C/C++ (e.g. Prolog, Lisp, Python, Kotlin combination).
Please, note that choosing a particular implementation platform might facilitate your system implementation or could make it harder. Please, do your research before you choose.

Step 3: Design an expert system
Initially it will involve knowledge formulation (just writing down some rules and conditions) and drafting some flow charts to indicate how it should operate.
In a case of a simple knowledge base, this stage blends from analysis into design and it is easily described by the creation of a matrix that lists some conditions along the top edge and recommendations on the side.

Step 4: Develop an expert system prototype.
This involves an actual expert system development with an application of the tool chosen or developed in step 2. For example, if you are using a shell, it will involve filling in this shell with some knowledge: simple rules and conditions and running it. After a prototype is created you have to test it, by running a number of consultations. Here you have to supply three-five examples. You can run your prototype on any platform in any environment. For example, if you develop the application that evaluates security of your Windows laptop environment, you can execute it on this platform and use any library (e.g. CLIPS). However, even if you design an expert system that will evaluate an Android device security, at this stage you can run an expert system on your Windows machine too.

Step 5: **Write documentation and submit a report.**

Submission Requirements for Project 2.

(1) Submission deadline is at 11.59 pm, on the date specified on MyCourses. Penalty for late submission will be 5% per day for individual projects and 10% per day for a group project. Note that the delay more than one hour will be qualified as a day.

(2) Please, upload your submission to MyCourses.

You have to submit:

a file named pr2_yourname(s).* which should contain your documentation in docx or pdf formats, including the user's guide and the program description. **PLEASE, submit this file separately. Do not include it into your archive.**
a file named pr2_yourname.**zip** which should contain all files zipped together. Please, submit ZIP archive only.
Depending on the project type, the contents could be different but the following files need to be present:
1) file named pr2_yourname.ext where ext is determined by the language you are going to use in your project, which should contain a commented code you developed for your project,

2) an executable file to run your expert system application,

3) other software files necessary to run your application.

This is your responsibility to make your submission sufficient enough to run your application with no extra information or specialized software necessary.

Documentation

Suggested report format

1. Executive summary (1-2 paragraphs)

Concise description of problem addressed and results

2. Requirements (1-2 paragraphs)

Your brief understanding of what the instructor requires in this project - informally stated

3. Specification (1 page)

Precise definition of what you are to do and what results you have to achieve. Please, be as specific as possible and provide as much detail as you can here.

4. Description of the domain problem

An overview of the tool (expert system shell), which you are using if you have chosen to apply a standard tool (about half of a page)

5. Feasibility study (1 page for group projects only). A feasibility study is an elemental study designated to decide and discuss various possibilities of the project parts implementations. It focuses on the project and outlines alternatives. The products of this study are used to make a decision whether or not to proceed with the project and which way. Note that this is a part of your report only. You do not have to implement all the discussed options.

Possible solutions, protocols, models and methods.

Comparisons (at least three comparisons of different options) of possible solutions (I want you compare at least three pairs of alternatives, e.g. Jess vs. Clips (this is just an example, please, use other ones)

Conclusion:

Choose one of each and provide your reasons.

6. Implementation (2 pages)

Representation of the knowledge base and data base

Description of the expert system applications (examples are required)

Structure, contents, user interface, limitations, software and hardware requirements, etc. Describe all the classes applied. If you used a code from some libraries, provide the reference to these libraries.

7. Testing description (1 page) – describe any tests you conducted and their results

8. User’s guide and GUI (if any) description (1 page)

9. Development process documentation (1-2 pages for group projects only): who was responsible for which part of the project and who was doing what

Course Project Part 3.

Use of machine learning techniques to implement your ES

Due: see MyCourses

This project consists of two parts. Part A is compulsory. It includes approximation of the ES input-output surface with a ML model. It could be implemented in any environment. You have to complete it. Part B is the bonus part and not compulsory. It has to be implemented on an Android device or Android simulator.

Part A.

This project is devoted to using machine learning techniques in your ES implementation. While an ES design has many advantages, its implementation requires a high computational power. Despite setting up such a goal, the developers of popular ES tools such as Jess, CLIPS, and SWI-Prolog have not yet produced a stable and reliable version for mobile devices. For example, authors of Jess ES engine announced an Android OS implementation version almost five years ago, but it is not available yet. We want you to investigate a novel approach to an ES implementation on mobile devices by utilizing machine learning (ML) techniques to approximate the hyper-surface that is produced by the ES.

Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions. Please, note that the following text specifies an employment of the artificial neural networks (ANN) and specifically their multilayer perceptron (MLP) model. However, you are encouraged to use other ML methods instead if you wish. In this case, you do not have to follow exact specifications below. But you have to produce a simple input-output relationship model with a good approximation of your ES that could be easily implemented on a resource constrained computation platform.

To initiate the supervised ML, you have to get a dataset first. In order to produce a dataset, you can generate all possible combinations (permutations) of all inputs and apply them to your ES model. To do this, you may iterate each input within a reasonable interval. Please, see sec. 3.4. of the attached paper for further details. Each data set should be divided into a training set and a testing set.

In ANN design, you have to research the choice of the number of neurons in a hidden layer. You will conduct your empirical study of different choices by computer simulation. In this study you will have to research the neurons number influence on the ANN performance and resource consumption. You may assume that we will use the three-layers structures with an input layer, a hidden layer and an output layer. The number of neurons in an input layer will equal to the number of inputs (or attributes applied for classification purposes). The number of neurons in an output layer could be chosen by different ways. You can see in Table 2 of the attached paper the ANN parameters that have been used but I want you to do your own research.

You will have to write and execute a code that will allow you to conduct your empirical study and process the results. You may use any programming language, any library, any design package to do this project but you have to provide a proper reference.

What do you have to in this project?

1. By feeding your ES prototype with various input values combinations and recording both inputs and outputs, you have to produce the dataset to be used in ANN training and testing. You have to change each input value within a certain interval in combinations with other inputs – see the attached paper for further details. After generating your dataset, you have to subdivide it into the training (around 67%), testing (and possibly) validation parts.

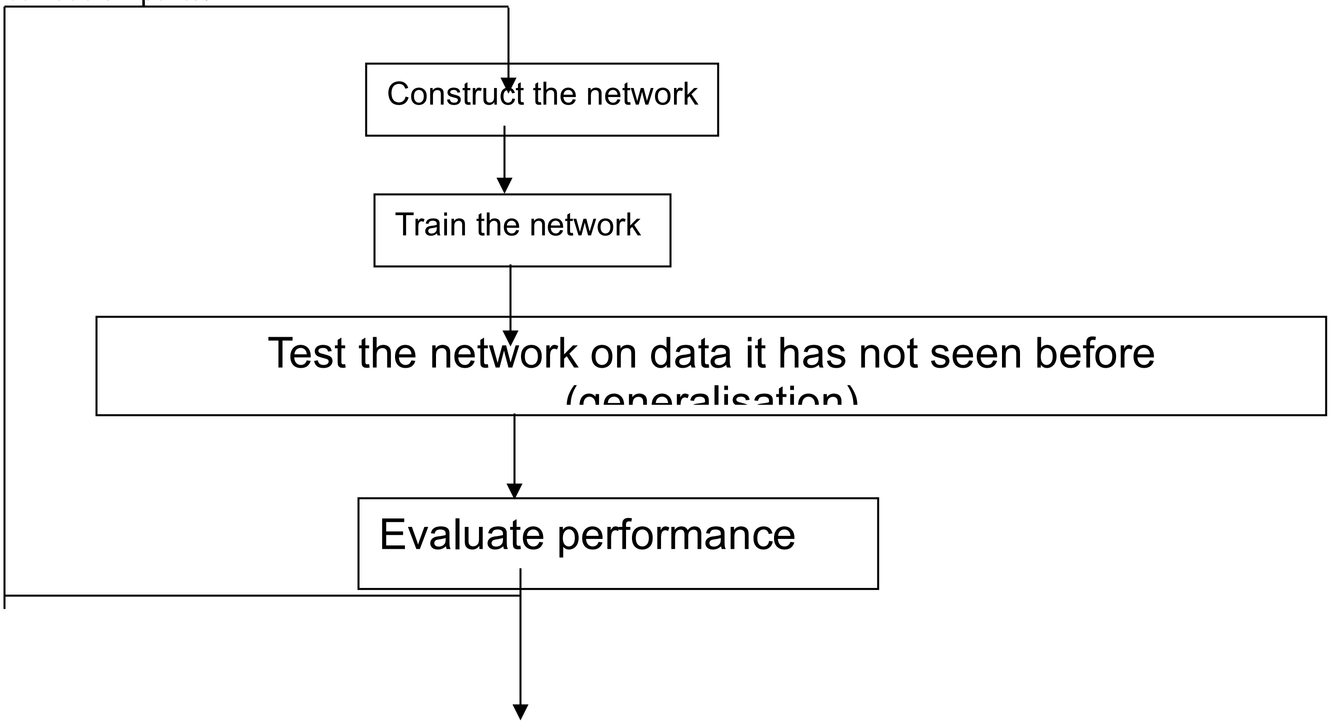
Construct the network
Train the network
Test the network on data it has not seen before (generalisation)
Evaluate performance

Fig.2. General ANN design algorithm

2. Choose the MLP architecture with an initial number of neurons in the hidden layer equal to about a half of a number of inputs, randomly generate weights and train your ANN with the goal to achieve a reasonable error on your training set. After the training is completed record the final error achieved and the time and memory consumed (time in terms of time units and epoch numbers).
3. Take the ANN produced by your training and apply your testing set to it. Record time and memory consumed for testing and the performance indicators (the error between ES surface and the generated ANN input-output surface – you can use the maximum or the average error or both of them).
4. Make the number of neurons in the hidden layer about twice as much as in the previous step, repeat steps 2-3.
5. Repeat step 4 a few times (at least 4). You have to get the error values pretty low (less than 1%). If negative, try to repeat step 4 a few more times. Now you will have the results indicating how the performance and resource consumption depend on the number of neurons.
6. Plot a few graphs demonstrating how the performance (approximation errors) and resource consumption depend on the number of neurons. (Hint: Matlab package is very good at producing various plots or you can use other tools).

Part B.

BONUS STEP: develop an app of an Android based smartphone that should evaluate a smartphone environment security.
You have to implement your expert system as a smartphone app. I should be able to execute it and get evaluation results. While executed, the app should provide an evaluation of the smartphone environment, in which it is run. You can use your own Android device or an Android emulator (https://developer.android.com/studio/run/emulator).
There are implementations of CLIPS that can be run on Android OS, however to use them you have to know Android NDK (Native Development Kit, which is different from Android SDK). Using Android NDK can be very time consuming task.
You also can write your own implementation or use third party expert system shells (JAVA implementation) that can be run on the Android OS. For example, you can design your project similar to https://github.com/bennapp/forwardBackwardChaining.
Please, note that **I am NOT teaching a smartphone app design** in this class. You have to use your existing knowledge and skills or do your own research.
You can find some applications of Android system security evaluation on the Google Play, for example:
https://play.google.com/store/apps/details?id=com.igorkh.trustcheck.securitycheck
https://play.google.com/store/apps/details?id=dataqualitylab.rit.ver_app_finder
In addition, if your application will satisfy Android OS guidelines (https://material.io/guidelines/material-design/introduction.html) we can assist you in publishing your application on the Google Play.

If you do this BONUS option, you are allowed to skip Test 1 in this class.

Please, let me know about your intention BEFORE the test.

Submission Requirements for Project 3.

(1) Submission deadline is 11.59 pm, on the date specified on MyCourses. Penalty for late submission will be 5% per day for individual projects and 10% per day for a group project. Note that the delay more than one hour will be qualified as a day.

(2) Please, upload your submission to MyCourses.

You have to submit:

a file named pr3_yourname(s).* should contain your documentation in Word or pdf formats.
a file named pr3_yourname(s).zip should contain your source code (including design packages like Matlab, if any).
Please, **zip** all your files together. Please, submit your report as a separate file in pdf or doc format. Please, do NOT use tar or other archives. Only zip files will be accepted.

Documentation

Suggested report format
Executive summary (1-2 paragraphs)
Concise description of the problem addressed and the results
Requirements (1-2 paragraphs)
Your brief understanding of what the instructor requires in this project - informally stated
Implementation (1 page)

Please, provide some explanation and user's guide for any code you designed and ran (e.g. structure, contents, user interface, limitations, software and hardware requirements, etc.) If you use a code from some library, provide the reference to this library.

Experiments (0.5-1 page)

Describe your experiments

Results (2-3 pages)

Describe your results, analyze and comment on them, please, include figures, plots and charts and describe them.

GOOD LUCK!

Attachment 1. Grading scheme

PROJECT GRADING SCHEME	Max points	Your points	comments
Project 1 paper:	**10**		
At least two-four (depending on your group size) papers reviewed and analyzed. The papers are not referenced in the papers attached.	3		
Conclusions are made regarding security evaluation ES and its implementation with ML	1		
Conclusions are clear and comprehensive	1		
Projects 2 and 3 specifications is produced	1		
Project specification is clear and workable	1		
Novel and interesting ideas of this project development are proposed (please, specify them in a separate section)	3		
Total:	**10**		
Project 2:	**20**		
able to compile/start and control execution	4		
Has a batch/makeup file to run or it is clear how to run it	1		
Gives a clear understanding how it works	1		
Number of rules is reasonable, at least 12 – 20 (could be smaller if this is mainly a programming project)	2		
Delivers a chain of rules: the output of one is the input to another	2		
Gives an understanding how the conclusion/output is produced	1		
Testing was reasonable, designed and implemented	1		
Documentation 2:	**8**		
Report has a reasonable size and is clear	2		
Feasibility study is present	1		
Feasibility study is reasonable	1		
User's guide is present	1		
User's guide could be followed and implemented	1		
Testing is described properly	2		
total	**20**		
Project 3	**20**		
ML architecture is used in conducting an empirical study with the required number of experiments investigating the influence of the neurons number on the performance with various numbers of neurons in a hidden layer	**10**		
Report includes a good description of your empirical study and the analysis of your results	**7**		
Report has a reasonable number of plots, charts and other figures	**3**		