PRE-PROJECT - REPORT

FOR BACHELOR THESIS

NTNU
Norwegian University of Science and Technology

TITTLE:
Machine learning and statistics for a predictive maintenance concept

CANDIDATES:
Magnus Gribbestad **Eivind Fugledal** **Kelvin Sundli** **Robin Bergseth Vågeskar**

DATE:	SUBJECT CODE:	SUBJECT:	DOCUMENT ACCESS:
11.01.17	**IE303612**	**Bachelor thesis**	- Open

CORSE/PROGRAM:	NO. PAGES/APPENDIX:	BIBL. NO:
AUTOMATION ENGINEERING	12/	- Not in use -

PRINCIPLE(S)/SUPERVISOR(S):
Bertil Helseth Ibrahim A. Hameed Ottar L. Osen

TASK/SUMMARY:

CONTENT

1 INTRODUCTION

Intelecy is a new start-up company with the idea to create a cloud-based self-learning predictive maintenance application for the process and manufacturing industry.

The student group have a general interest for data processing and development of AI based algorithms. This led to an agreement between Intelecy and the students that the focus on the thesis is going to be pre-processing of data and the research and development of intelligent algorithms.

The reason these topics are interesting is that the user could potentially reduce a considerable amount of waste by being able to predict that systems are about to fail or break. Reductions in downtime, failures, and materials will result in both reduction of cost and improvement of the users' environmental footprint. The user will have a more reliable process which produces a higher quality, due to increased understanding of their processes.

In addition, variations of artificial intelligence are a rapidly developed and utilised, which makes this an opportunity to follow the development of a modern way to process data and optimized solutions.

2 NOTATION

- Predictive analytics
  - A variety of techniques to analyse historical and current data, to predict the future events.

- Artificial intelligence, AI
  - A machine doing a task that would require intelligence if done by a human.

- Data pre-processing, pre-processing of data
  Decide how to deal with missing, out-of-range, redundant and impossible combinations of data.
  - Cleaning of data. Inputting values, where missing.
  - Normalization of data. Setting a known range to the data.
  - Transformation of data.
  - Feature extraction. Extract relevant features from the data. Could be frequency or skewness.
  - Selection. Which data do we need? Which data is usable?

- Intelligent algorithm
  - An algorithm that produces answers to a task that would require intelligence if done by a human.

3 PROJECT ORGANIZATION

3.1 Project group

Student number	Name	Role
279372	Magnus Gribbestad	Project manager
267448	Eivind Fugledal	Secretary pt. 3
266079	Robin Bergseth Vågeskar	Secretary pt. 2
266111	Kelvin Sundli	Secretary pt. 1

Table 1: Project group and positions

3.1.1 Project group tasks - organization

Everyone in the group has an equal responsibility of carrying out the project. Each member is responsible for keeping project manager up to date on progress and deviation from plan.

3.1.2 Project manager tasks

Keep Gantt-diagram up to date.
Preside at meetings with group of supervisors.

3.1.3 Secretary tasks

- Draw up the minutes of meetings.
- Call meetings.
- Reserve rooms for meeting.
- Write report of progress.

3.1.4 Tasks for group members

- Assist project manager and secretary

3.2 Supervisors

Name	Organization
Bertil Helseth	Intelecy
Ibrahim A. Hameed	NTNU
Ottar L. Osen	NTNU

Table 2: Supervisors and their organizations

4 AGREEMENTS

4.1 Agreement with client

The project is to develop possible methods for data pre-processing and predicting from data. These data are at first simulated data provided by Intelecy. Real data from potential customers might be used at a later point. The group will try to find dynamical, adaptable, and modular solution designs in the form of techniques, rules, and algorithms. The work shall be done in a scientific manner.

See attached agreements, per NTNU standard.

4.2 Workplace and resources

The project is to be carried out in NTNUs campus in Ålesund.

In addition to Bertil, Ibrahim and Ottar as resources, the group can also contact Leif Larsen and Espen Davidsen at Intelecy for assistance. Leif is a software developer with knowledge of Microsoft Cognitive Services, and Espen has a master's degree in distributed information systems.

Intelecy will supply licenses for Azure and provide courses in Azure Machine Learning and Schneider Wonderware.

Essential parts of the results, made from Intelecy's basis, might be withheld from public access for two years, per attached agreements. This is to avoid revelation of techniques, rules and algorithms that might be essential to Intelecy's existence.

4.3 Agreements of the group

The group is agreeing on having a core time, where every member needs to attend. The core time is set to be from 09:00 to 14:00, Monday to Friday. Exceptions can be made if a group member needs to be away, and **notices** the rest of the group as soon as possible.

Meetings is to be arranged every second week, preferably before 12:00 on Mondays. Exceptions may occur if group is working on other courses or the supervisors are unavailable.

Disagreements regarding the project resolution are to be voted internally by all members of the group. If number of votes are equal, the last vote is to be given by Ibrahim.

Disagreements regarding the group's work method is to be determined by voting or 'best of three – rock paper scissors' between disagreeing parts.

5 PROJECT DESCRIPTION

5.1 Problem - target – purpose

The problem can be divided into several parts

- How can we analyse existing sensor data to predict future problems, break-downs, and needed maintenance?
- How can we recognise anomalies and create signatures of known failures?
- How can we pre-process the data to make it usable and reasonable for the analysis tools?

Value target:
We want to create a prototype from the concept. This prototype should be able to predict future issues with an acceptable accuracy, for a user to have economic and environmental savings.

Result target:
The concept should, if possible, form a working prototype for pre-processing and predicting from data. The prototype should, with at least 80% accuracy, predict

failures. Our desire is that the prototypes should be easily adapted and applied to real-world problems.

This prototype should be built on techniques for data processing and predictivity that the group creates together.

Process target:
Learn to work in project teams and organizations.

Utilize previously acquired knowledge and get new knowledge about data processing and predictivity.

5.2 Project specification

We are creating a system for predictive maintenance. Our focus is the pre-processing of the sensor data, and the models for predictivity.

This system needs to predict issues from the data with an accuracy that adds value for the user or operation. The system should predict the yield failures with at least 80% accuracy. This means that each prediction (fail or pass) should have an 80% likelihood of being correct.

5.3 Development procedure

The progress estimations will be adapted dynamically with a Gantt-diagram, due to insufficient basis now. The diagram will be made with focus on continuous adaptability.

Each group member will coordinate a pre-defined set of tasks. One member is coordinating each task, but several members will be involved in the tasks due to priority and focus on learning outcomes. The coordinator of each task is fully responsible for the implementation of his task.

The project manager should be notified as soon as a problem occurs, so that a meeting can be arranged, to find an appropriate solution.

5.4 Data collection

After delivery of the pre-project, the group will get hold of and study relevant literature. The relevant literature is divided into several sub-parts, for which each group member is responsible for collecting the information. These responsibilities are specified in the Gantt-diagram.

In addition to study relevant literature, each member is also responsible for exploring existing work within the field of predictive analysis and pre-processing of data. Existing work may give a head start to our work.

All members have previously used MATLAB for data processing and principal testing and three out of four members of the group have taken the course 'Intelligent Systems' which is a good foundation for the analysis.

5.5 Risk management

The target is that we should have a working prototype for predictive analysis.

Reasons to succeed:

1. Well composed and organized team
2. Well defined tasks and responsibilities
3. Good teamwork
4. Good communication internally in group and between group and supervisors

Reasons to fail (risks):

5. Short on time
6. Bad communication
7. Bad teamwork
8. Vague task distribution plan
9. Unusable data
10. Too much time needed to learn tools and principals
11. Absent members (due to sickness or other reasons)
12. Signals give insufficient information about critical anomalies

		Consequence			
	Risk management	Low	Medium	High	Very high
Probability	Very high				
	High				12
	Medium			5, 10	
	Low		9	6, 7, 8, 11	

After classifying our risks and their consequences we see that most of the factors are within what is "good". Risk no. 5 and 10 can be reconsidered at a later point, when knowledge is sufficient.

Risk 12 – *Signals give insufficient information about critical anomalies*. The probability of this is high, and the impact would be very high. To avoid this, we need to have simulators that represents real life applications well. If the available sensor data from the target processes are insufficient, we must consider generating rules for a minimum set of sensors or even supply additional sensors that is giving enough information. This might be sonic, ultrasound or vibration sensors.

The risk management matrix shows that our risks generally have a low probability of happening and have low impact on our project. This will allow us, with quite high certainty, to complete the project.

If we find a solution that works as specified, it might have a huge positive effect on the users' economy. In such case, it will also provide the users with opportunity to improve their environmental footprint by optimizing their maintenance routines and spare parts storage.

5.6 Main activities

NO.	MAIN ACTIVITY	RESPONIBLE	COST (NOK)	TIME/ SCOPE (DAYS)
A1	Pre-project report	KS	0	3
A2	Specification of concept	MG, EF	0	6
A3	Literature study – Research	All	0	17
A4	Application framework	KS	0	6
A5	Data processing	All	0	18
A6	Machine learning	KS	0	21
A7	Analysis and evaluation	EF, RBV	0	35
A8	Conclude	RBV	0	16
A9	Report	All	0	98
ALL	Project total		0	~98

These are the main activities. Please refer to the updated Gantt-diagram for better explanations and sub-tasks.

5.7 Progress plan

5.7.1 Main plan

The project will proceed with respect to the Gantt-diagram. The Gantt-diagram includes estimated date of start and approximate number of days for each sub-task. The diagram also specifies the coordinator of each task, who is responsible for the implementation of the respective task.

For detailed plan, please see Gantt-diagram.

Milestones	
Description	Within date
First actions in Azure	27.01
Available data processing rules collected	03.02
Available predicting models collected	03.02
Every group member knows how to build an easy predictive system in Azure	06.02
First data pre-processing setup is ready	24.02
First model for signalling anomalies is ready	02.03
First test of predictive system is done	07.03
Conclusion regarding best results is reached – Specify accuracy and precision	08.05

5.7.2 Project coordination tools

To manage our project, we have started by using the tool GanttProject. This tool lets us specify main activities, sub-tasks, coordinators, and time scope. Milestones are given in the table above.

To follow software development progress, we are going to use Visual Studio Team Service for issue and task tracking. For collaboration and document sharing we are using SharePoint and ShareLaTeX.

5.7.3 Development tools

The platform for running the actual pre-processing and machine learning algorithms is Microsoft Azure. Visual Studio Online and bitbucket is used for code hosting. In addition, we will use MATLAB and PyCharm to test solutions and ideas, and comparing them.

5.7.4 Internal evaluation

The internal control in this project is performed at the morning meetings. Here any issues should be detected and easily corrected. The morning meetings will also update the group members on the progress.

For an objective to be counted as complete the group must agree that it is at least good enough to be used in any subsequent objectives. This means that code is checked and has no errors or warnings, documentation is updated and all members of the group agrees that it is sufficient.

For a milestone to be counted as complete all the objectives of the milestone have to be fulfilled.

5.8 Decision-making process

During the time spent working on the pre-project, the group as a unit have taken the initiative to a bachelor thesis from Intelecy and discussed what aspect of the product that Intelecy is developing, the group found most interesting. The group have together specified the content of the thesis from what the group found most interesting. Every decision made have been discussed with the whole group present.

For every major decision, the group must make during the project, the group must arrange a meeting where the decision is discussed. The decision shall be presented to the supervisors before the decision is put into effect.

6 DOCUMENTATION

6.1 Reports and technical document

At the end of the project, the group are going to deliver a scientific main report with a technical description of the result and methods, and appendix containing pre-project report, code, minutes of meetings, meeting request, test results and progress report.

6.1.1 Formulation

All documents shall be written in English and follow predefined templates in ShareLaTeX or MS Word. All documents shall follow predefined writing rules.

The main report shall follow the NTNU standard template for bachelor thesis.

Code, produced by the group, shall use camelback syntax with English words, and have describing comments.

Meeting request shall be distributed by e-mail to the members of the group and supervisors.

6.1.2 Content

The main report shall contain a technical description of the results and methods.

Meeting requests shall contain date, time, location, meeting agenda and progress report.

The minutes of meetings shall contain the key points of the meeting, the agreements made by the participants, and the participants attended.

The agenda shall contain the point to be discussed in the meeting, and have a miscellaneous point at the end for inputs during the meeting not directly related to the points in the agenda.

6.2 Routines

All documents shall be backed up to a shared folder in SharePoint, Google Drive or in ShareLaTeX, and all code shall use a distributed version control system. Backup of documents and code shall be done at least once a day.

Minutes of meetings shall be written and distributed to the group members and supervisors the same day as the meeting.

Group members shall report to project manager once a week with updates on the status of the tasks.

Meeting requests shall be distributed at least 24 hours before meeting.

Group meeting every morning at 9am.

6.3 Storage

Documents and code are stored locally on the group members' computers and remotely online. Online storages are SharePoint, Google Drive, ShareLaTeX and Bitbucket.

7 PLANNED MEETINGS AND REPORTS

7.1 Meetings

7.1.1 Meetings with supervisors

Meeting with the supervisors shall be held every 2 weeks. The meeting is preferred to be at the same day every time. Monday is the preferred meeting day.

The meetings shall follow the agenda, and inputs not directly relevant to points on the agenda shall be collected and discussed at the end of the meeting in the miscellaneous point of the agenda.

The purpose of the meetings with the control group is to inform the supervisors on the progress of the project and for the supervisors to guide the project group on the progress.

The secretary is responsible for taking notes during the meeting and writing the minutes of meetings and distribute this as soon as possible.

7.1.2 Project meetings

The project group will have a short meeting every day at 9.am to update each other on their progress and the plan of the day.

7.2 Periodic reports

7.2.1 Progress report

Before every meeting with the supervisors, a progress report shall be made showing the planed and performed tasks in the last two-week period. The progress report shall describe possible changes or deviation from the plan and the planned activities for the next two weeks. In addition to the progress report, the progress plan (Gantt-diagram) shall be updated before each meeting with the supervisors.

8 TREATMENT OF NONCONFORMANCE

If a deviation of a task occurs, that the person responsible for the activity cannot solve, the project group must be notified. In such case, the project group needs to arrange an internal meeting where an evaluation of the deviation shall be made.

These questions shall be asked:

- Can the deviation be solved with more resources?
- Can the deviation be avoided?
- Can the deviation be solved with external help?
- Can the task be reformulated so that the deviation is avoided?

The supervisors shall be notified after the project group have evaluated the deviation. This gives the supervisors the opportunity to give feedback before a final decision is made.

9 EQUIPMENT

The group´s need for equipment is only software related, and every group member shall have access to the following software tools:

- MS Azure
- Wonderware
- Word
- ShareLaTeX
- MATLAB
- PyCharm
- Gantt-project
- Visual Studio Online
- Bitbucket
- SharePoint

The group's prerequisites for implementation is for the principal to supply data sets for training and testing of algorithms and systems.

ATTACHMENTS

Attachment 1	Gantt-diagram
Attachment 2	Confidentiality agreement
Attachment 3	Standard agreement