

CONTENTS

1.	Defining the purpose of the thesis	5
1.1.	Preliminary definition of Data Science.....	5
1.2.	Acquisition of measurement data for analysis	6
1.3.	Methodology for the analysis.....	7
2.	Identifying problems that arise with the collection of an increasing amount of measurement data.....	9
2.1.	Performance problems that arise with large data sets	9
2.2.	Reusable analyses for similar datasets.....	9
2.3.	Functional limitations	9
3.	Searching for available methods of data analysis already used in mechanical engineering....	11
3.1.	Excel (Microsoft).....	11
3.2.	Origin (OriginLab)	11
3.3.	LabVIEW (NI)	12
3.4.	Solartron Metrology.....	13
4.	Presentation of the general outline of data science methods and the possibility of using them to analyse measurements.....	15
4.1.	Data science	15
4.2.	Lifecycle of data science project	15
4.2.1.	Understanding problem	15
4.2.2.	Data collection.....	15
4.2.3.	Data cleaning and processing	15
4.2.4.	Exploratory data analysis	15
4.2.5.	Model building and evaluation.....	16
4.2.6.	Communicating model results	16
4.2.7.	Model deployment and maintenance	16
4.3.	Programming language (Python).....	16
4.4.	Tools	16
4.5.	Use cases	16
5.	Measurements of plain bearings at available test stands.....	17
5.1.	Measurement stand	17
5.2.	Measuring sleeve.....	18
6.	Use of data science methods to analyse obtained data from plain bearing measurements....	19
7.	Development of the analysis results and conclusions.....	21
	Bibliography	23

LIST OF FIGURES

1.1	Data Science graph.	6
3.1	Excel GUI	11
3.2	Origin GUI	12
3.3	LabVIEW GUI	12
5.1	Measurement stand	17
5.2	Example of measurement sleeve	18

1. DEFINING THE PURPOSE OF THE THESIS

The main objective of this paper is an attempt to use the tools provided by Data Science to analyse data from measurements made for research work in mechanical engineering.

In some applications, a conventional data analysis methods for mechanical engineering are beginning to be insufficient and a limitation to obtaining the best possible conclusions from measurements. Here it would be useful to look around at what data analysis capabilities have been developed in recent years and try to implement them for mechanical engineering. The leading branch of science focusing on data analysis in recent years is actually Data Science and it seems that using the tools that have been developed there can bring a lot of benefits to processing measurements data.

Data science methods are not limited by the existing software - one can freely shape the interface and use any statistical tools. Data science tools based on python and R libraries are optimised for processing large databases, use of machine learning methods is increasingly used to make predictions based on previous experiences.

Thanks to the fact that most of the libraries around Python and R are open source and built by the community, they very often respond directly to the needs of users and are constantly extended and new ones appear. More information about programming languages and available tools will be presented in Chapter 4.

1.1. *Preliminary definition of Data Science*

To begin with, it seems appropriate to give an overview of what Data Science is, the origin of this branch of science and what branches of science it is related to. The following diagram, presented below on figure 1.1, can be useful for this purpose; it shows the overlap of the branches of science that data science combines.

One can then recall a short definition, the unfolding of which, combined with the diagram above, should give a fairly good intuition about the complexity of this domain.

"Data science combines the scientific method, math and statistics, specialised programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data." [4]

First of all Data Science is the domain of study that deals with vast volumes of data which are collected from different sources. Here we are dealing with the lowest circle in the diagram 1.1, the data can come from a business that wants to analyse its actions (for example, it can be traffic data from a store's website, or GPS data from trucks of a transport company) or from specialised research from which we want to draw the greatest insight. The data used for analysis can be from multiple sources and present in various formats.

Data Science uses modern tools and techniques that come from both circles on the top of graph 1.1. Techniques derived from mathematics and statistics provide long-established capabilities for computing, analysing, and processing data. Computer science, on the other hand, provides the ability to perform these calculations on large data sets and curate machine learning algorithms to build predictive models.

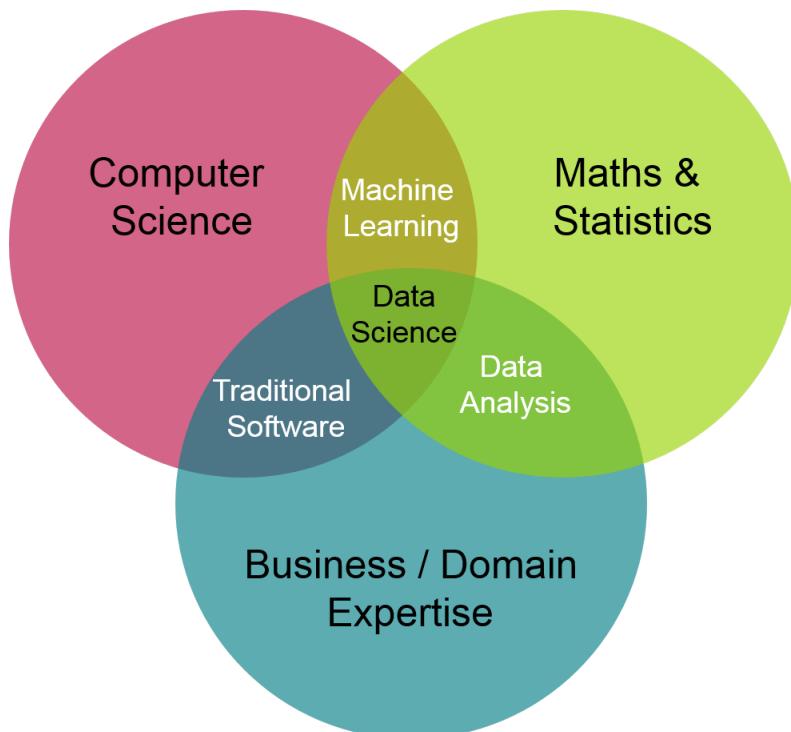


Figure 1.1: Data Science graph. [3]

The combination of all this gives the ability to find unseen patterns, extract meaningful information and make business decisions that would not be possible with conventional solutions.

The above short definition shows that it is a field of science which takes a multidisciplinary approach and is aimed at the most accessible presentation of data in a business environment where it is often necessary to show the results of some actions to people who are not involved in a given field. Using these methods focuses on getting the most utility out of data, which in today's world is a large and often growing set.

This short introduction will be developed in one of the following chapters (Chapter 4) where aspects such as: lifecycles, tools, use cases and more will be presented in depth.

1.2. *Acquisition of measurement data for analysis*

Thanks to the possibility of cooperation with the department of mechanical engineering, it was possible to obtain data resulting from the work on plain bearings that they conduct. Initial discussions revealed a need to expand the range of tools for analysing the data obtained from this research for reasons discussed in more depth in Chapter 2 of this thesis. Thanks to this possibility, it was possible to obtain real data from measurements and there was no need to create measurements or even simulate what measurements would be just to complete this paper.

Another advantage will be the possibility of comparing the results of analyses based on data science methods to analyses conducted by a unit that deals with these bearings on a daily basis, and possibly supporting their work with tools that can be selected in further stages of this work. In the case of the team that deals with these measurements, they have already been in contact with data scientists from outside the university in the form of orders to create software for processing data from specific measurements. From this point of view, it seems all the more reasonable to

combine efforts to explore the topic of data analysis using data science methods.

1.3. Methodology for the analysis

The preliminary idea of the work presented in this paper is to analyse the results of measurements using selected data science methods and to compare these results with the results of analysis performed using conventional methods of mechanical measurement analysis that have been used so far by the team who perform the measurements.

2. IDENTIFYING PROBLEMS THAT ARISE WITH THE COLLECTION OF AN INCREASING AMOUNT OF MEASUREMENT DATA

Improvements in the quality of measuring instruments and increase in their precision allow us to perform measurements with increasing accuracy and resolution. Thus, the amount of data that we obtain from measurements is also growing and it seems that this trend will continue in the coming years. Hence, there is a need to increase the possibility of processing these data, preferably in a complete free way and giving the flexibility to choose tools for their analysis from the widest possible spectrum. In such a way as to be able to adopt to the incoming new data in case of a desire to extend measurements or change the measuring equipment and thus add new output results or possibility of adopting a different data structure.

Data processing methods, which have been used so far in the process of measurement in mechanical engineering starts to limit possibilities of extracting the greatest possible amount of informations from the measurements.

2.1. *Performance problems that arise with large data sets*

For measurements with the latest high-resolution sensors, a single measurement can consist of several thousand measurement points or more. Spreadsheet-type graphical programs start to be difficult to use in such cases, and problems may occur from the very moment of viewing the data. Further, there are limited possibilities of data cleaning as well as limitations with presenting data in the form of graphs, especially when it comes to graphs presenting three-dimensional measurements.

2.2. *Reusable analyses for similar datasets*

If we consider the case in which we want to measure a large number of similar elements - in this case sleeves that are sliding bearings, there arises a desire to have a tool that will enable automatic start of analysis on successive sets of data from measurements.

With conventional programs, this requires duplicating a spreadsheet and pasting more data there. In case of a script written for a given type of measurement it is enough to run it on another file and almost immediately we get desired graphs or any other form of results.

2.3. *Functional limitations*

3. SEARCHING FOR AVAILABLE METHODS OF DATA ANALYSIS ALREADY USED IN MECHANICAL ENGINEERING.

When looking for a reference point, it is necessary to trace the software used so far for analysing the measurement data. In order to narrow down the spectrum, the focus has been on the software used to date to process data from plain bearings measurements which will later be analysed using data science methods.

3.1. Excel (Microsoft)

The first thing that comes to mind when analysing data is, of course, Excel. The most well-known, easy-to-use and available software that is used for all kinds of analysis, processing and viewing data. The interface of the program is shown in Figure 3.1.

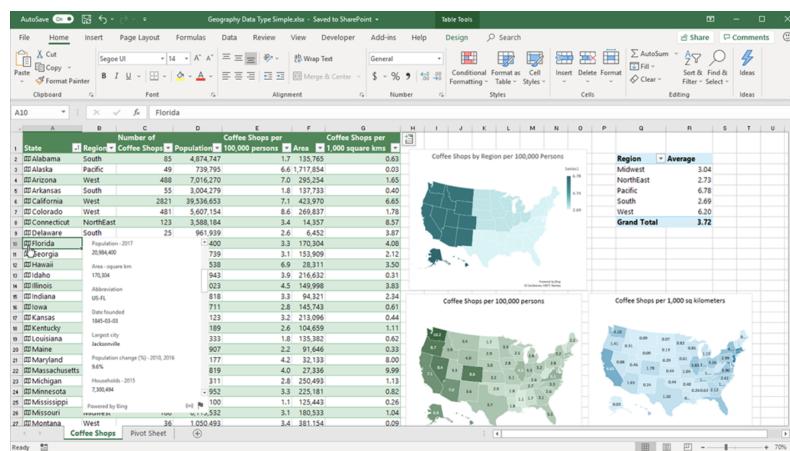


Figure 3.1: GUI of Excel software. [5]

"Microsoft Excel provides a grid interface to organise nearly any type of information. The power of Excel lies in its flexibility to define the layout and structure of the information you want to manage. Basic tasks require no special training, and Excel allows you to work with text, numbers, and date information in a relatively open and unstructured way. Nearly 30 years after its initial introduction, Excel remains the world's leading spreadsheet software." [6]

3.2. Origin (OriginLab)

Origin is GUI based software aimed at scientists and therefore has a lot of capabilities often needed for measurement analysis. The software has a lot of available options, add-ons and possibilities to work with other programs. Graphic user interface of Origin software is shown on Figure 3.2.

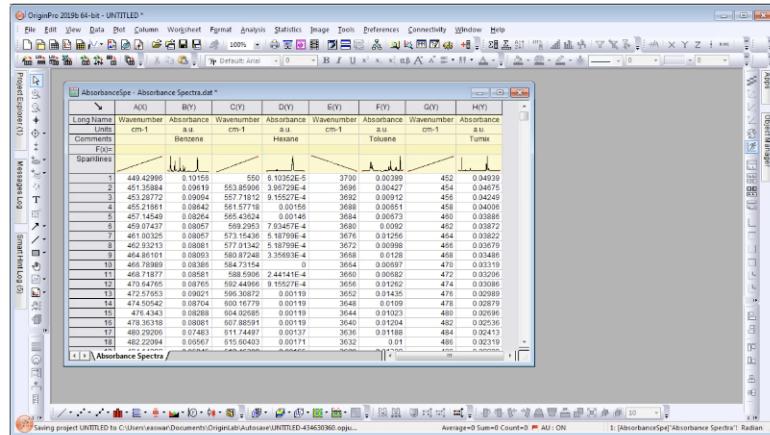


Figure 3.2: Origin GUI. [7]

"Origin is the data analysis and graphing software of choice for over half a million scientists and engineers in commercial industries, academia, and government laboratories worldwide. Origin offers an easy-to-use interface for beginners, combined with the ability to perform advanced customisation as you become more familiar with the application. Origin graphs and analysis results can automatically update on data or parameter change, allowing you to create templates for repetitive tasks or to perform batch operations from the user interface, without the need for programming. Extend the capabilities in Origin by installing free Apps available from our website." [7]

3.3. LabVIEW (NI)

LabVIEW is used to create a measurement strategy, build a test bench, and quickly view real-time data. Its capabilities are quite broad and it is a good test environment because of the ease of making changes and access to simple visualisation methods immediately after measurements. An example of the LabVIEW graphical interface can be seen in Figure 3.3

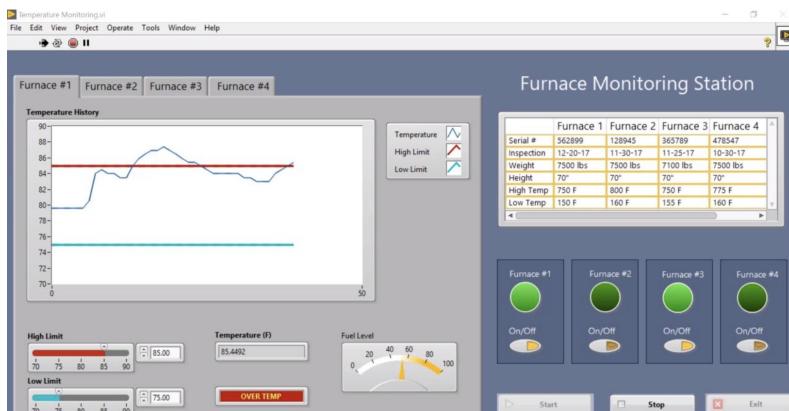


Figure 3.3: LabVIEW GUI. [8]

"LabVIEW is systems engineering software for applications that require test, measurement, and control with rapid access to hardware and data insights. LabVIEW offers

a graphical programming approach that helps you visualise every aspect of your application, including hardware configuration, measurement data, and debugging. This visualisation makes it simple to integrate measurement hardware from any vendor, represent complex logic on the diagram, develop data analysis algorithms, and design custom engineering user interfaces." [8]

3.4. Solartron Metrology

In this particular case, one of the software that will be used in this research work will be a detector program from Solartron Metrology. Due to the measuring devices used, the program is needed to process the output signals from the measuring devices. The signal that arrives at the computer is then processed by the software and prepared for further analysis in the form of files such as .cvs. There is a simple interface here both to control the sampling method and to adjust the format of the output data. In the following discussion the role of the software from Solartron will be omitted because the focus is on the analysis of the data itself and not on the way of conducting the measurements.

4. PRESENTATION OF THE GENERAL OUTLINE OF DATA SCIENCE METHODS AND THE POSSIBILITY OF USING THEM TO ANALYSE MEASUREMENTS.

4.1. *Data science*

4.2. *Lifecycle of data science project*

When it comes to how a project is run in data science, there is a term called project lifecycle. This may differ in the way it is described as well as in the specific points depending on the source where you look for information about the project lifecycle, but in principle the main steps will be similar:

4.2.1. *Understanding problem*

Even before first insight into the data it is necessary to define and understand the problem which is a subject of analysis. Deep understanding of the problem principles is the most important in early stage because it can protect from loosing time for development of model which won't meet assumptions. In data science most of the projects is based on problems brought in from business or scientific demand and hence it is very important that the communication between those aggregating data and those doing strictly data science analysis is clear. Data analysis itself should not be in isolation from those doing the measuring, and there needs to be a constant exchange of insights and guidance on the desired output.

4.2.2. *Data collection*

Once you understand the problem, it is important to collect the right data sets. Data is the core of the whole problem and it is necessary to collect it in an appropriate way to be able to create a project based on it. In the case of this particular paper, the data is collected by researchers working in the field of plain bearings on a daily basis and on specially designed for these purposes test stands.

4.2.3. *Data cleaning and processing*

It is not enough to just collect data, because most of the time such raw data will not be suitable to create a model. Once collected, the data needs to be properly processed and cleaned before it can be used for the rest of the work. It is necessary to solve flaws and inconsistencies as well as deal with null values and outliers.

4.2.4. *Exploratory data analysis*

At this stage of the project there is already data as well as it has been cleaned and pre-processed and organised. One can take a deeper look at this data to get familiar with the properties and features. There is also an occasion to build a so-called intuition about the data set and

begin to think about a strategy for creating the model. After a closer look at the data, there is an opportunity to raise a deeper discussion with experts in the field, try to find the first patterns and visualisations.

4.2.5. Model building and evaluation

Although the previous phases of the project take a lot of time and are important, it is only during the building of the model that the most important things happen. The first thing is to properly divide the database into train and test set. The test set is used to evaluate the model and check whether it works correctly only on the dataset that was used to build it. This is especially concern in case of predictive models based on machine learning where in case of too small train data set or too many learning cycles performed on this set model can lose flexibility and be useless after loading new data set.

Various metrics provided by the statistical domain can be used for evaluation. The evaluation methods will depend on the specific problem, and after the first feedback, adjustments to the model can be made.

4.2.6. Communicating model results

4.2.7. Model deployment and maintenance

4.3. Programming language (Python)

In this particular paper, the programming language used for data analysis will be Python, which main advantages are versatility, simple syntax, multitude of available libraries, large user base, ever-growing popularity in the scientific community, and much more.

"Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed." [9]

4.4. Tools

4.5. Use cases

5. MEASUREMENTS OF PLAIN BEARINGS AT AVAILABLE TEST STANDS.

The main measurement will focus on trying to analyse the loss of material in the inner surface of the sleeve which works as a plain bearing. The surface of the sleeve is asymmetrically eroded and the main task in analysing this measurement will be to determine the volume of material loss.

5.1. *Measurement stand*

For this purpose a measuring stand has been constructed as shown in Figure 5.1. The stand consists of:

- control panel - used to determine the method of measurement, among others the angle of rotation of the measured sleeve, the task of displacement of the sensor for subsequent angular measurements
- rotary table - the measured sleeve is fixed to it, driven by a belt drive
- sensor - attached to the vertical axis, built on a trapezoidal screw
- other mechanical and electronic parts - drives, housings, control system, cable interface

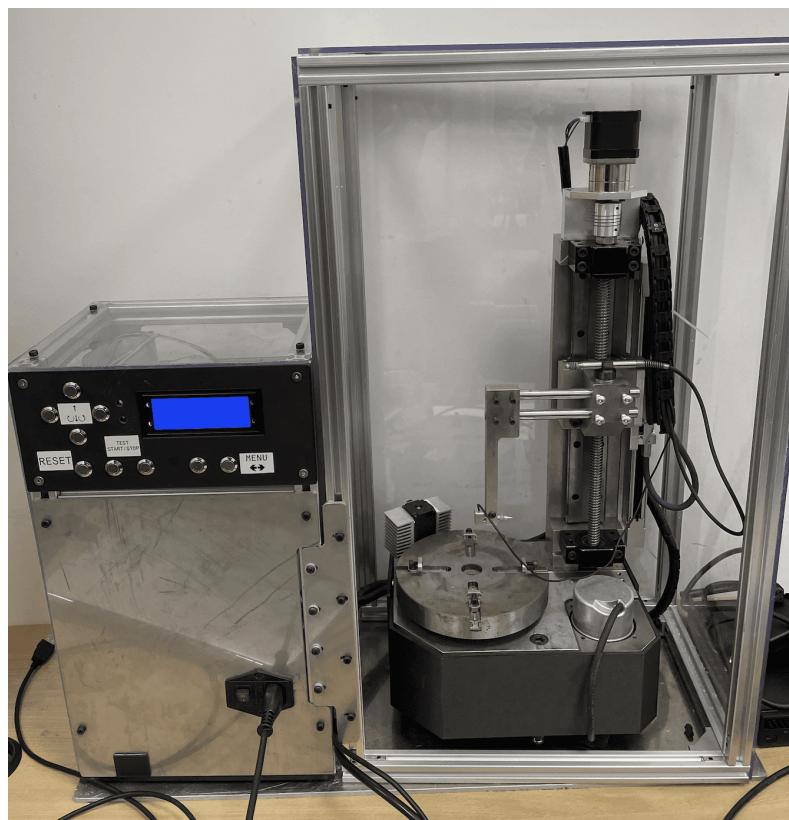


Figure 5.1: Measurement stand.

When it comes to the measuring station itself, it is worth mentioning that its inaccuracy must be taken into account. The measurement accuracy of the sensor exceeds the precision with which the position of the sensor can be set. Z-axis built on a trapezoidal screw with the driving motor has

a height error of 0.1mm (?), while the rotary table has a position accuracy of 0.1 degree (?) and there is also the error of making the rotary plate of 0.05mm (?). While analysing the measurement results, there will also be an attempt to find these bugs and neutralise them if possible.

The data that are received from the measuring station are prepared by the software attached to the sensors themselves in a specific way. The data is immediately converted into Cartesian coordinates and additionally the information about the workstation settings for a particular measurement is added. For analysis, data are retrieved from .csv files, a format supported by both spreadsheet programs and easy to load into Python.

5.2. Measuring sleeve

Figure 5.2 shows an example of the sleeve that will be measured, one can see that the sleeve consists of two materials - an outer metal material and an inner plastic layer.



Figure 5.2: Example of measurement sleeve.

A significant loss of inner layer material on one side of the sleeve can be observed. The object of the analysis will be to determine as precisely as possible the volume of this loss based on measurements taken at several heights of the sleeve.

**6. USE OF DATA SCIENCE METHODS TO ANALYSE OBTAINED DATA FROM
PLAIN BEARING MEASUREMENTS.**

7. DEVELOPMENT OF THE ANALYSIS RESULTS AND CONCLUSIONS.

BIBLIOGRAPHY

- [1] A. Geron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, Inc., 2017
- [2] P. Bruce, A. Bruce, *Practical Statistics for Data Scientists*, O'Reilly Media, Inc. 2017
- [3] <https://thedatascientist.com/data-science-without-programming/>
- [4] <https://towardsdatascience.com/data-science-101-life-cycle-of-a-data-science-project-86cbc4a2f7f0>
- [5] <https://www.microsoft.com/en-us/microsoft-365/blog/2019/02/06/excel-with-microsoft-excel-in-office-365/>
- [6] <https://www.opengatesw.net/ms-excel-tutorials/>
- [7] <https://www.originlab.com/index.aspx?go=PRODUCTS/Origin>
- [8] <https://www.ni.com/pl-pl/shop/labview.html>
- [9] <https://www.python.org/doc/essays/blurb/>