isep Instituto Superior de Engenharia do Porto

Many Labs

Application User Manual

João Moreira 1190709

Jorge Ferreira 1201564

Rafael Leite 1201566

Rui Pina 1201568

Santiago Azevedo 1201623

GLOSSARY

- E-mail - Electronic mail (email or e-mail) is a method of exchanging messages ("mail") between people using electronic devices.
- Password - a secret word or phrase that must be used to gain admission to a place.
- JavaFX - JavaFX is a software platform for creating and delivering desktop applications, as well as rich web applications that can run across a wide variety of devices.
- DoubleClick – With you mouse use de left button and double click to make an action.
- Report - give a spoken or written account of something that one has observed, heard, done, or investigated.
- Diagnosis - the identification of the nature of an illness or other problem by examination of the symptoms.
- Test - take measures to check the quality, performance, or reliability of (something), especially before putting it into widespread use or practice.
- NHS - The National Health Service (*NHS*) is the umbrella term for the publicly funded healthcare systems of the United Kingdom (UK).
- Sample - a small part or quantity intended to show what the whole is like.
- Test Results – information about the outcome of a laboratory analysis and does not include personal identifying information about the subject.
- Employee – a person employed for wages or salary, especially at non-executive level.
- Company – a commercial business.
- CPU – A central processing unit (CPU), also called a central processor, main processor or just processor, is the electronic circuitry that executes instructions comprising a computer program.
- Clinical Analysis Laboratory – is a laboratory where tests are done on clinical specimens in order to get information about the health of a patient as pertaining to the diagnosis, treatment, and prevention of disease.
- RAM – Random-access memory (RAM) is a form of computer memory that can be read and changed in any order, typically used to store working data and machine code.
- TIN – is a unique combination of letters or numbers assigned by a jurisdiction to an individual or an Entity and used to identify the individual or Entity for the purposes of administering the tax laws of suck jurisdiction.

USER MANUAL

TABLE OF CONTENTS

INTRODUCTION

Purpose and Scope

This User Manual contains all essential information for the user to make full use of the Many Labs app. This manual includes a description of the system features and capabilities, as well as annexes with math content. It is addressed to every user who interacts with this company and that need to use this application.

System Overview

Many Labs is an English company that has a network of clinical analysis laboratories and that as an application to manage the clinical analyses performed in its laboratories.

This application allows workers to register data like tests, samples, reports and more, according to their function on the company. Clients can also enter the app, to check tests or update personal data.

Every user can access the application with their e-mail and password, defined in their registration.

One of the goals of Many Labs with this application is to provide a modernized platform, intuitive and easy to use, common to all users.

(core modules and diagram)?

System Requirements

The application requires a machine with (minimum) 8GB of RAM, and needs an intermediate CPU usage, to support the use of JavaFx and some minor calculations. The application should run on all platforms for which there exists a Java Virtual Machine

Software Installation

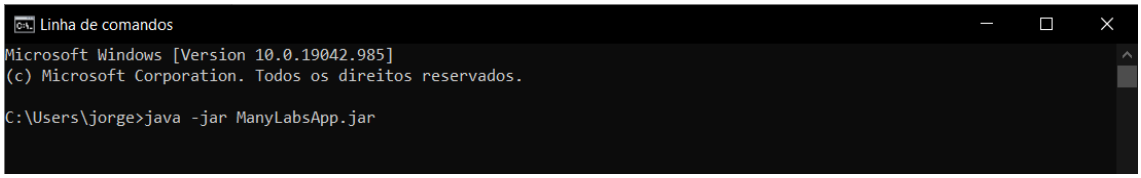
For using this app you will need to execute a jar file in your command line. After enter in the directory where the file is, all you need to do is write java -jar ManyLabsApp.jar, and then the application will start.

SYSTEM FEATURES

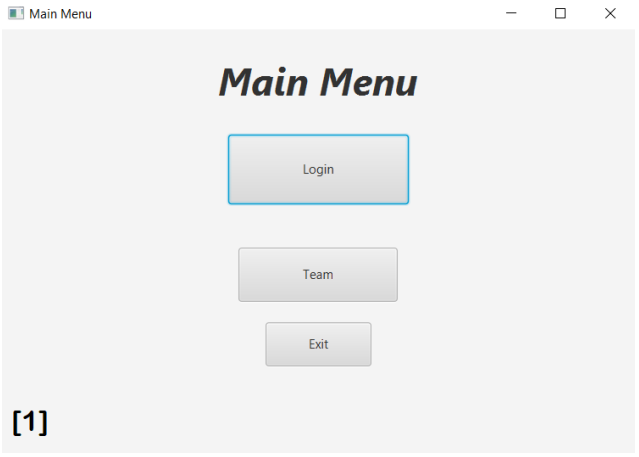
LOGIN

When logging in, you will be allowed to perform the tasks that are defined, according to your role in Many Labs.

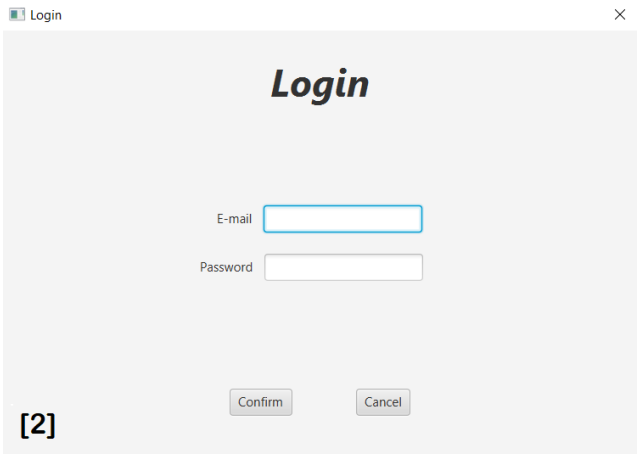
1) Execute the jar file in your system command line and press **ENTER**. Make sure that you are in the same directory of the file.



2) After pressing **ENTER**, a window will open [1]. Then you select the button Login, and another window will open, with spaces to write your email and password [2].



[1]



[2]

3) Type your **E-mail** and **Password**.

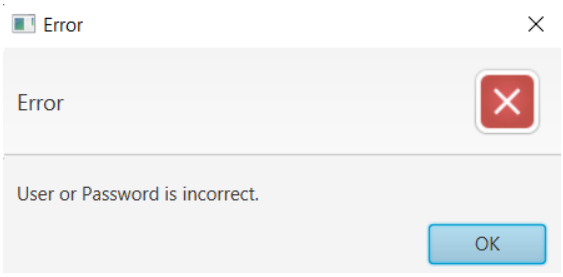
E-mail needs to have the character '@' and the domain at the end.

Password is an alphanumeric code with 10 characters.

4) Click **Confirm** button.

5) **Successful Login**. A window with your role features will be open. This is what it looks like if you are a laboratory coordinator or a clinical chemistry technologist, but there are other roles.

6) **Unsuccessful Login**. This window appears. You need to verify if you have entered your credentials correctly. Click button OK or X and try to log in again.

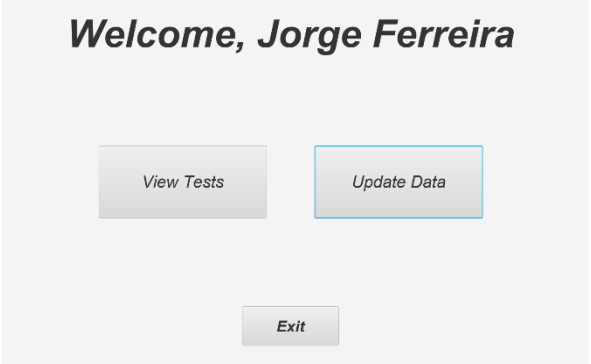


7) There is not a feature to change your password, so please be careful and make sure that you do not forget it.

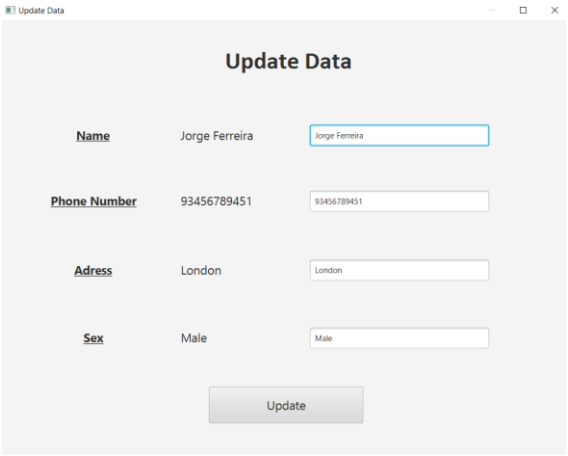
Client Features

Update Client Data

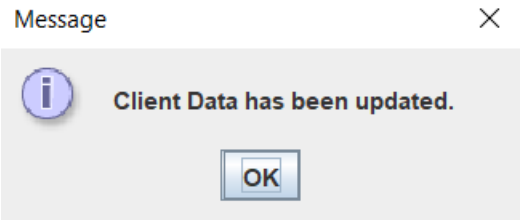
1) After logging in, a window with your name will appear. Then, you need to click the button **Update Data**. (In this example, the client is called Jorge Ferreira).



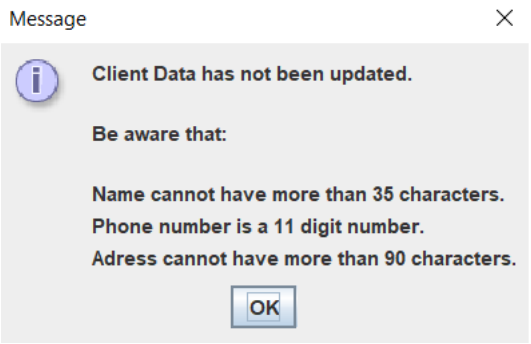
2) Update your data, by filling the blank boxes, and then click the button **Update**.



3) If you enter the data correctly, a message like this will appear and you will also receive an email, confirming that your data has been updated.



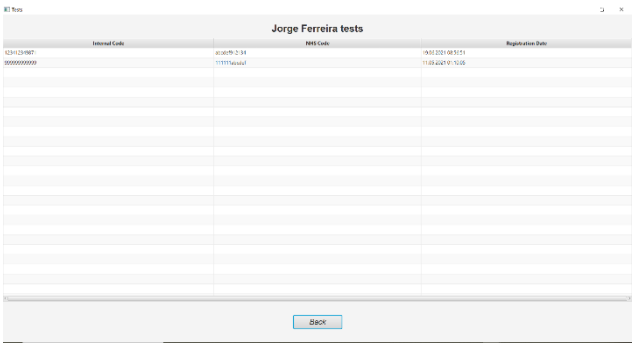
4) If you entered the data incorrectly, a message like this will appear, remembering you the rules to update your data correctly. Click button OK or X and try again.



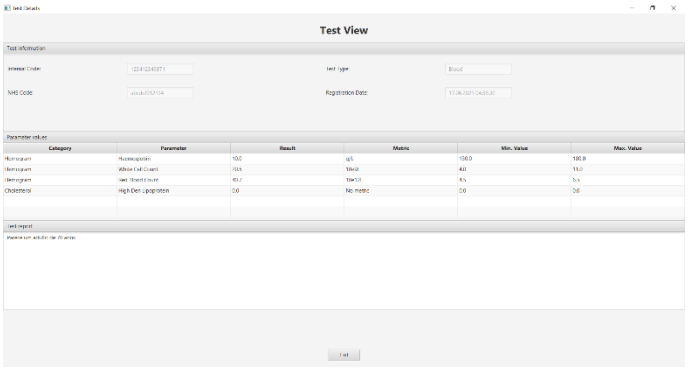
View Tests

1) After logging in, a window with your name will appear. Then, you need to click the button **View Tests**.

2) You will see a table with all your tests. To see the details, you can make a **double click** in the one you want or click on one test and then select **Test Details**.



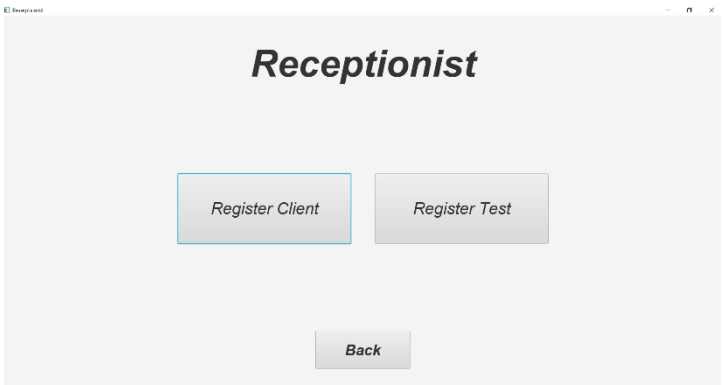
3) Done. Now the test details are displayed. You can see the parameters of the test, results, metric, references values and test information.



Receptionist Features

Register Test

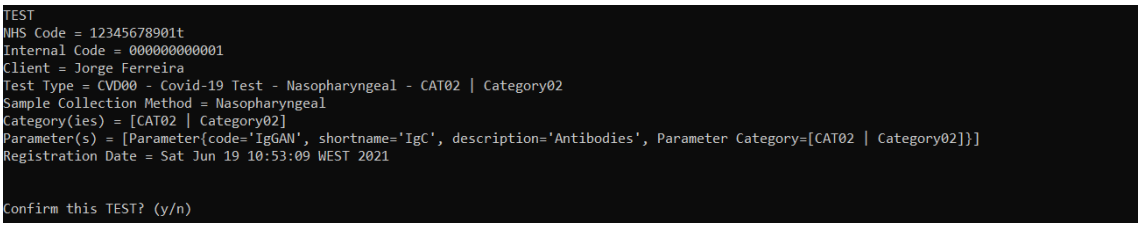
1) After logging in, a window to choose the Lab where you work will appear. Then, you need to click the button **Register Test**.



2) This feature works only on your cmd. You need to insert test information:

   - NHS code has 12 alphanumeric characters and is unique.
   - If the client is not registered in the system, you need to register him before insert TIN.
   - After, select one test type and then, for every category chosen, select the parameters that are in client's lab order.

3) Confirm the operation (y/n). y – Test saved. n – Test not saved.



Register Client

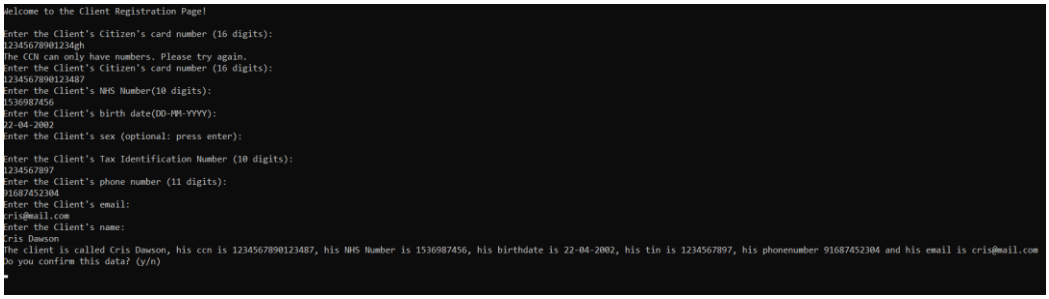
1) After logging in, a window to choose the Lab where you work will appear. Then, you need to click the button **Register Client.**

2) This feature works only on your cmd. You need to insert client information:

   - Citizen Card Number is a 16-digit number.
   - NHS number is a 10-digit number.
   - TIN is a 10-digit number.
   - Birthday needs to be type in DD/MM/YY format.
   - Sex - Male/Female. This field is optional.
   - Phone number is a 11-digit number.
   - The name has at maximum 35 characters.
   - E-mail needs to have the character '@' and the domain.

3) Then all the client data will appear on the screen, and you need to confirm the work. (y – Client saved. n – Client not saved).

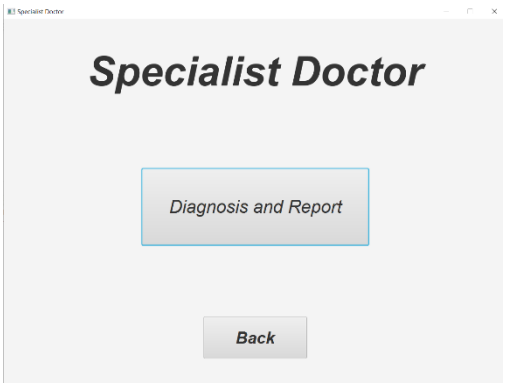
4) The client will receive an e-mail with the password to login into the application.



Specialist Doctor Feature

Write Report

1) First you need to log in accordance with this role. Then you need to select the button **Diagnosis and Report**.



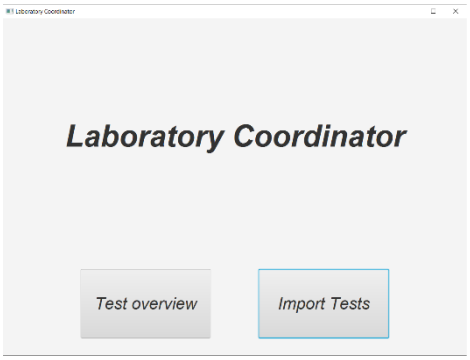
2) This feature works only on cmd. A list of tests ready and validate will appear and you need to select one.

3) After selecting one, the results will appear, and all you need to do is write the report and confirm the work.

Laboratory Coordinator Features

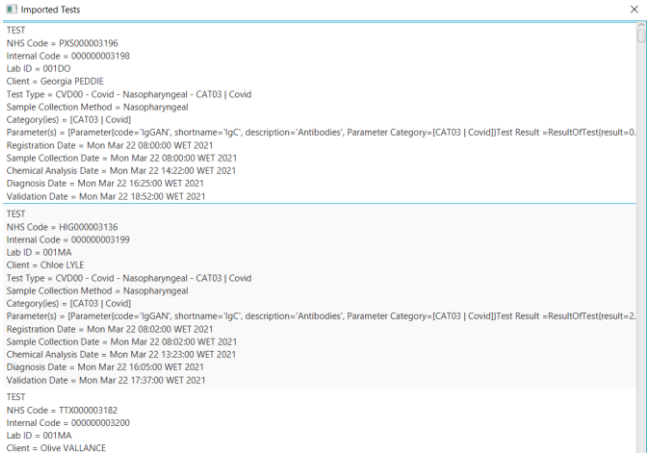
Import Tests

1) Enter in the app, do the login, and select the button **Import Tests**.



2) Select a CSV File in your computer, that is in accordance with the format that the company works with.

3) A window will open with all the tests imported from the file that you picked.

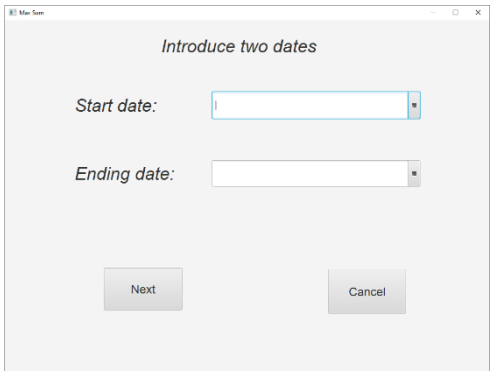


4) All the lines where a test was not imported, and the reason why, will be shown in the console.

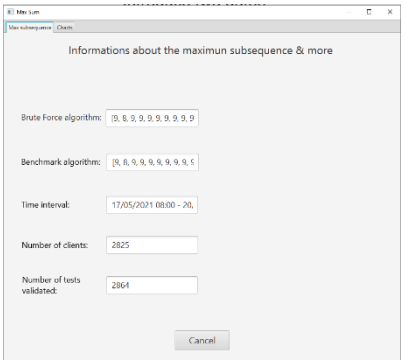
Test overview

1) Enter in the app, do the login, and select the button **Test Overview**.

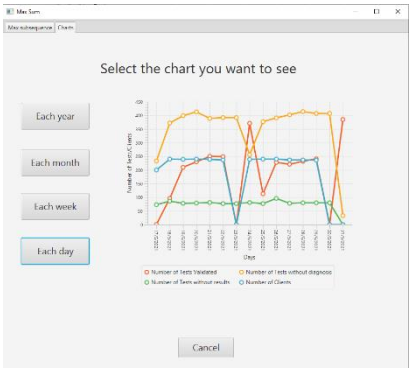
2) Select two dates that will serve as the company's study range and click “Next”.



3) A window will open with a variety of relevant information about the company's performance.

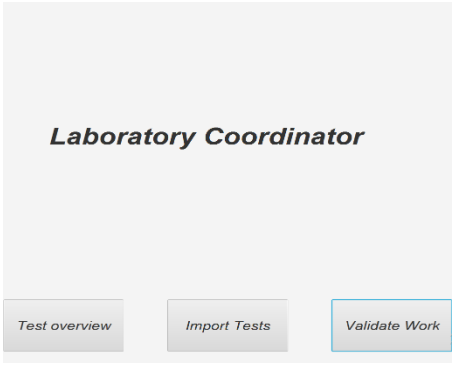


4) In this same window you can analyze charts created from the dates previously entered.



Validate Work

1) Enter in the app, do the login, and select the button **Validate Work**.



2) After that the application will open the console and show you a list of tests with a number before them that can be validated.

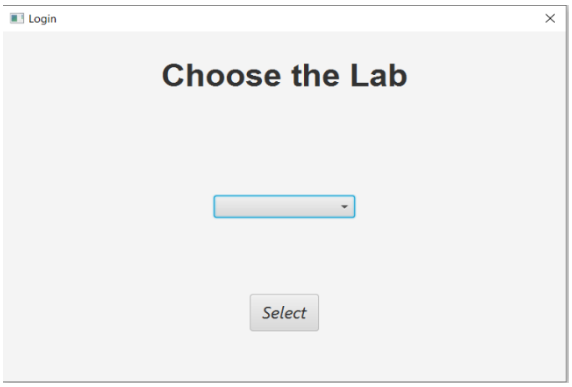
3) Choose the tests you wish to validate by typing the number before them separated by a space bar (or by typing the word “ALL” you can validate them all at once).

4) The application shows you the tests that are about to be validated and asks for confirmation. You confirm the validation depending on what you want.

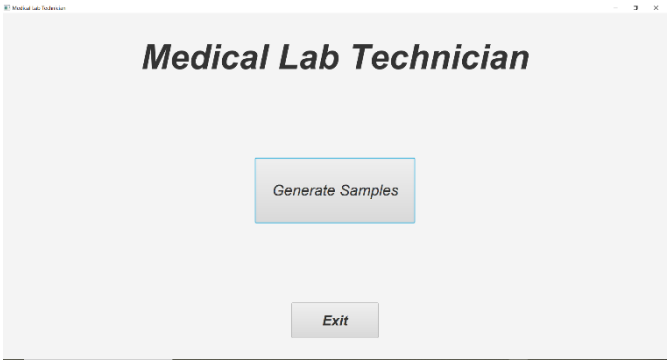
Medical Lab Technician Feature

Generate Samples

1) Enter in the app, do the login, and you will need to choose the Laboratory you work for.

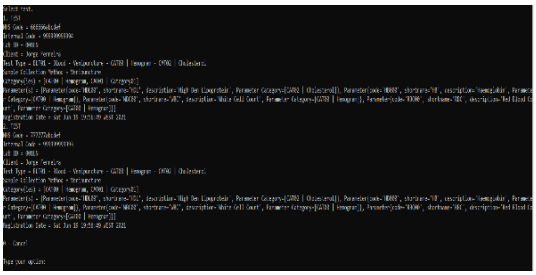


2) After selecting the Laboratory, the Medical Laboratory Technician menu will appear, and you need to select the button **Generate Samples**.



3) This feature only works in the cmd. You will need to choose the test you want to generate de samples and the number of samples you want to generate.

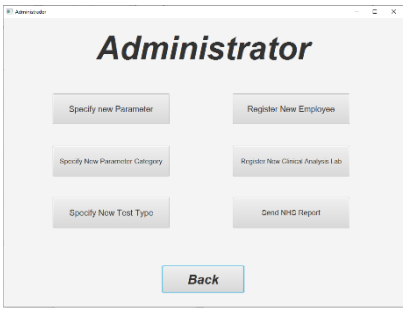
4) At the end it informs that the operation was successful and returns to the menu of the Medical Laboratory Technician.



Admin Features

Register an employee

1) Enter in the app, do the login, and select the button **Register New Employee**.



2) After that the application will open the console and give you instructions on what to type to create the employee. Follow those instructions.

3) When you are finished with typing all the attributes a confirmation message with the employee you are about to create will appear. Confirm their addition to the system or not if you want.

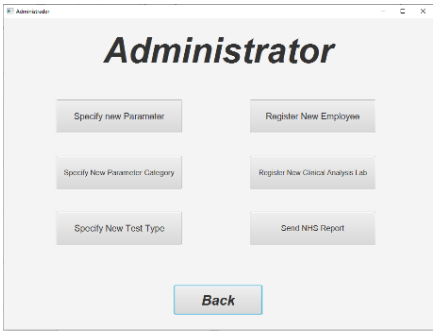
```
This employee is named Employee Name. Their ID is EN00001. Their address is Employee Adress. Their phone number is 123456789.
Their SOC is 1234. Their email address is employee@lab.com. Their username is employee1. Their password is 81LYGrj. Their role is RECEPTIONIST.
Do you wish to add the Employee you've registered? (Y/N)
```

4) If all the data, you typed was valid the employee will be added to the system. If not, the employee will not be added. The employee will also not be added if you answer negatively to the confirmation.

5) The system tells you the result of the operation and sends you back to the menu.

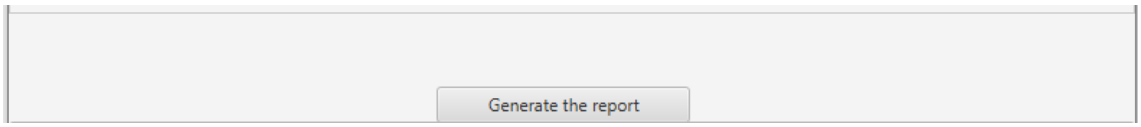
Send NHS Report

1) Enter in the app, do the login, and select the button **Send NHS Report**.



2) After that, a window will open, displaying the type of linear regression you want to use. Choose one of them.

3) Another window opens, this time, asking you for the data to generate the model, type the data accordingly (the significance may only be from a number from 0-1 and the parameters should be a number. Use the date pickers to pick the date. After all the data has been typed press the “Generate the report” button.



4) The report you are about to send will appear in the text field below. If the results are as you want them to be, you can press the yes button to send the report. If not, you can re-edit the data and type the “Generate the report” button again.

5) If you want to go back to the menu press the “X” in the top right of the application window.

Specify New Parameter

1) Enter in the app, do the login, and select the button **Specify New Parameter**.

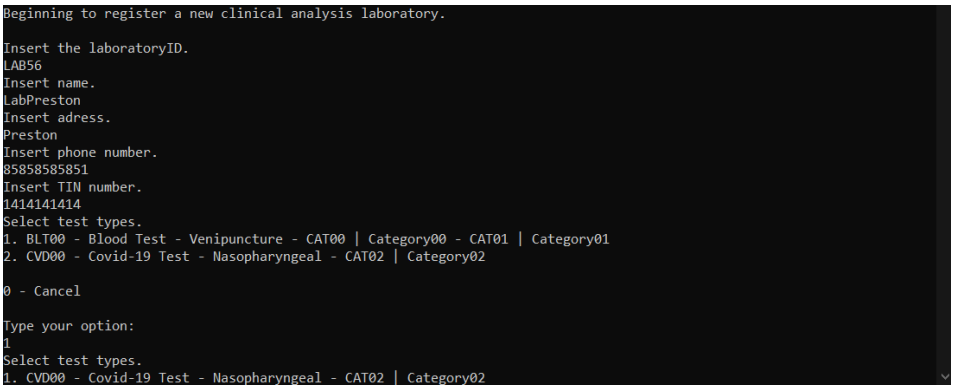
2) This feature works only on your cmd. Now you need to type the parameter data and select a category.

- Code - Five alphanumeric characters.
- Short Name - Max eight characters.
- Description - Max twenty characters.

3) Confirm the work (y/n).

Register Clinical Analysis Laboratory

1) Enter in the app, do the login, and select the button **Register New Clinical Analysis Lab**.



2) This feature works only on your cmd. Now you need to type the Clinical Analysis Laboratory data.

- Laboratory ID - five alphanumeric characters.
- Name – Max 20 characters.
- Phone Number - 11-digit number.
- TIN - 10-digit number.
- Two Clinical Analysis Laboratories cannot have the same data (except the name).

3) Confirm the work (y/n).

Specify New Parameter Category

1) Enter in the app, do the login, and select the button **Specify New Parameter**.

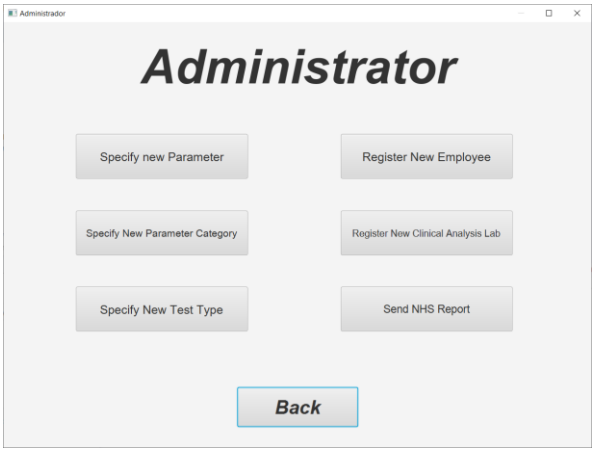
2) This feature works only on your cmd. Now you need to type the parameter category data.

- Code - Five alphanumeric characters.
- Name - Max ten characters.

3) Confirm parameter category creation (y/n).

Specify New Test Type

1) Enter in the app, do the login as an administrator, and select the button **Specify New Test Type**.



2) This feature only works in the cmd. You will start to specify a new test type. First is need to insert a code. This code can only have 5 digits and it needs to be alphanumeric.

```
Beginning to register a new type of Test.

Insert the code of the test type.
```

3) Then you need to insert a description. The description cannot be empty and has, at maximum, 15 chars.

```
Insert the description of the test type.
```

4) Next, you need to insert the collecting method. Again the collecting method cannot be empty but has, at maximum, 20 chars.

```
Insert the collecting method of the test type.
```

5) Finally, you can choose the parameter categories you want to associate to the test type. This cannot be empty.

```
Select parameter category.
1. CAT00 | Hemogram
2. CAT01 | Category01
3. CAT03 | Covid
4. CAT02 | Cholesterol

0 - Cancel

Type your option:
```

6) In the end, it will show all the data and ask you if you confirm to create or not.

Clinical Chemistry Technologist Features

Record Test Results

1) Make login and in the CCT menu select the button **Record Test Results**.

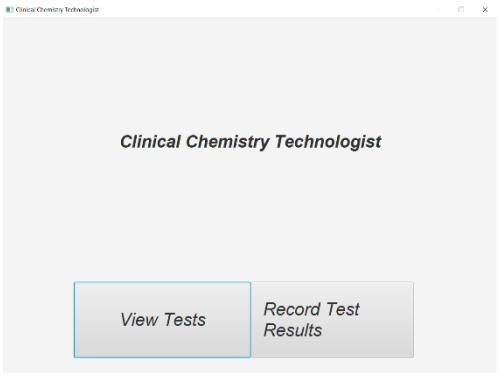
2) This feature works only on your cmd. You need to insert the barcode and then add the result of every parameter of the test.

```
Welcome to the Record Results of Test Page

Please introduce the barcode of the Sample where the results will be recorded.(11 digits)
11111111111
Select parameter to record the result.
1. Parameter{code='HB000', shortname='HB', description='Haemoglobin', Parameter Category=[CAT00 | Category00]}
2. Parameter{code='WBC00', shortname='WBC', description='White Cell Count', Parameter Category=[CAT00 | Category00]}
3. Parameter{code='PLT00', shortname='PLT', description='Platelet Count', Parameter Category=[CAT00 | Category00]}
4. Parameter{code='MCV00', shortname='MCV', description='Mean Cell Volume', Parameter Category=[CAT01 | Category01]}

0 - Cancel

Type your option:
1
Please insert the result for the intended parameter:
25
```

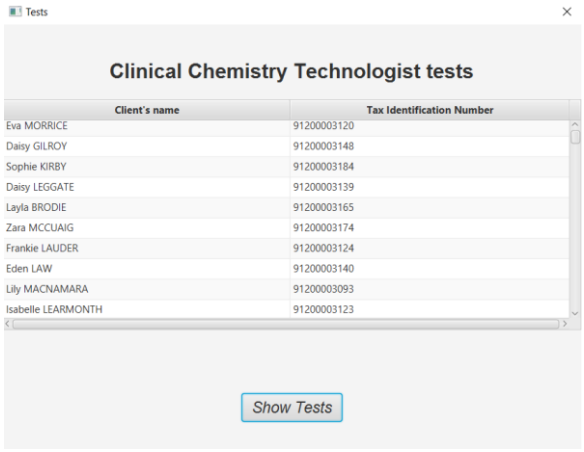
3) Confirm the operation (y/n).

View Tests

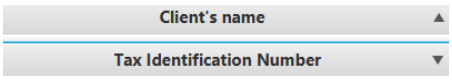
1) Make login and in the CCT menu select the button **View Tests**.



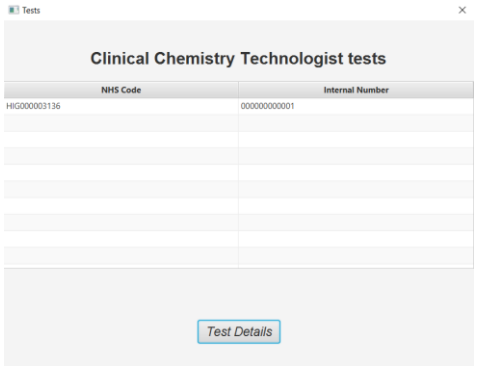
2) A window like this will appear. You got a table with the Client's name and their Tax Identification Number, and you need to select the one that you want to see the test. Click on the line and then click on **Show Tests**.



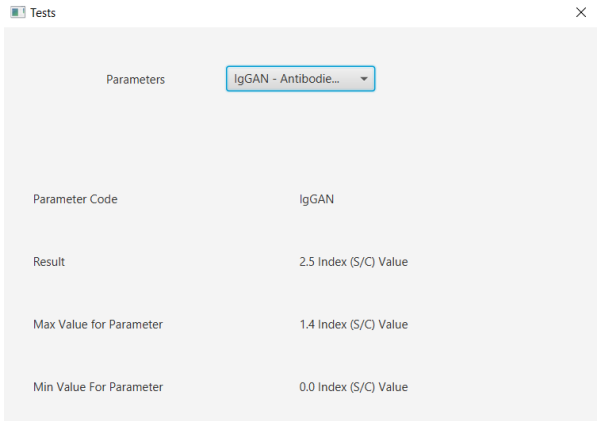
3) You can order the Client's names and the TIN's by clicking in the header. The arrow tells you if you are in ascending or descending order.



4) Now select the test that you want to see the results.



5) Now you can see the test result for every parameter. Select the parameter in the box and then the result will be shown in the screen.



System Feature

There is an additional feature which does not require a human to operate it. The system automatically sends a report to the NHS every day at 6 am. The report to be sent however needs to be configured prior by the administrator in a configuration file.

```
Automatic.Report.CurrentDate = 31-05-2021
Automatic.Report.TypeOfDate = days
Automatic.Report.HistoricalPoints = 30
Automatic.Report.DateInterval = 12-04-2021 24-05-2021
Automatic.Report.TypeOfLinearRegression = Multilinear
Automatic.Report.ConfidenceValue = 0.95
```

With these parameters the system creates the report and sends to the NHS whenever it is 6 am.

Troubleshooting

Issue #1: As an administrator I am pressing the “Generate Report” button, but nothing appears.

Solution: That probably means that you have typed something that is not a number in the fields where there should only be numbers.

Issue #2: The report to be sent to the NHS only contains 0s and Nans. Why is it so?

Solution: If your report only contains 0s and Nans means that there were no positive covid tests performed.

Issue #3: The report to be sent to the NHS only contains 0s and Nans. Why is it so?

Solution: If your report only contains 0s and Nans means that there were no positive covid tests performed on the interval you have selected.

Issue #4: I have the following message on my console: "You must insert an interval of dates superior to 2 days/weeks.". What does this mean.

Solution: When doing the calculations for the Report it is needed to calculate the degrees of freedom which is a parameter that cannot be 0 or negative. This parameter is calculated according to the number days/weeks you introduce to your date intervals. For this not to happen just make sure the interval is between two days/weeks (e.g 14/04/2021 18/04/2021).

Issue #5: I wanted to generate samples to a particular test and it does not show that test. Why is it so?

Solution: If the test you want to generate samples for are not showing it means that probably it does not in the laboratory you are working for or there was an error with the registration of the test.

Issue #6: I want to see my tests and they do not show up. What should I do?

Solution: The application will get all your validated tests, if there are no tests showing up it means that your tests are not validated yet. You will receive a notification via SMS and Email when it is validated.

Issue #7: I am creating a test type and in the end there is an error that does not save. What is the problem?

Solution: Probably the code that you have entered already exists so you will need to create again with a different code

Issue #8: I am logging in and it shows an error when I try to login. What is the problem?

Solution: The application does not accept the email and password empty, and if the login does not exist it will appear the error too. Check again the data you are inputting and try again.

João Moreira 1190709 - 1190709@isep.ipp.pt Jorge Ferreira 1201564 - 1201564@isep.ipp.pt

Rafael Leite 1201566 - 1201566@isep.ipp.pt Rui Pina 1201568 - 1201568@isep.ippt.pt

Santiago Azevedo 1201623 - 1201623@isep.ipp.pt

FAQs

Q – How can I add a new TestType?

A – To add a new TestType you can easily do that if you are the Administrator. For this just access the Specify new Test Type functionality which is described here.

Q – Why can't I see all the information about a client?

A – The only role who has access to all the information is the Specialist Doctor and therefore if you are not one you can access it.

Q- How can I switch the algorithm to be used in the sorting of clients/tins?

A – This can be easily changed in the configuration file which comes with the application. Just locate the comment "#Algorithm to sort Clients by Name and Tin" and change the value beneath to any other Algorithm you want to be used. It is to note that the Vanilla Application only features 2 sorting algorithms and other can be added.

Q – I want to add a result to a test; however, I cannot see it. Why is it so?

A – If you cannot see the test you want to add a result there are two possible explanations. Either the system has an undetected bug (which is more unlikely) or the test has not any samples associated to it. We recommend you check if the test you want to add a result has gone through the process of generating samples. If it has gone through that process, please contact us via the contact information.

Q – I want to add a result to a test; however, I cannot see it. Why is it so?

A – If you cannot see the test you want to add a result there are two possible explanations. Either the system has an undetected bug (which is more unlikely) or the test has not any samples associated to it. We recommend you check if the test you want to add a result has gone through the process of generating samples. If it has gone through that process, please contact us via the contact information (which is located on the end of this section).

Q – As a client can I change my password? No, you cannot. Please refer to the functionalities described in here

A - As a client can I change my e-mail? No, you cannot neither for the same reason as this question.

Q - In the Import Test feature, can I choose any CSV file with tests?

A - No. The file needs to have the formats accepted by the company.

Q - What is the application key?

A - The key to the application is to manage information about blood and covid tests.

Q - Can i change the language in the application?

A - No. The application only supports English. Since it is only intended for use in the UK and is linked to the NHS services.

Q - Can I change the language in the application?

A - No. The application only supports English. Since it is only intended for use in the UK and is linked to the NHS services.

Q- If I create something in the application will it be saved?

A - Yes, but only if you exit the application. If the application breaks and you do not exit it will not save.

Q- If I create something in the application will it be saved?

A - Yes, but only if you exit the application. If the application breaks and you do not exit it will not save.

Q – Where will the barcodes be saved?

A - In the folder of the application there is a folder named "Barcodes" and it is saved inside it.

Q – Where will the barcodes be saved?

A - In the folder of the application there is a folder named "Barcodes" and it is saved inside it.

Q - How do I restart the application data?

A - In the application folder there is a folder entitled "data.dat". Just delete it and it should restart all the application data.

Q - Who is responsible for all the test data?

A - The company that you are associated with.

Annex A MATCP

1.1 Overview of Simple Linear Regression

Simple linear regression is used to estimate the relationship between two quantitative variables. You can use simple linear regression when you want to know:

How strong the relationship is between two variables (e.g. number of positive Covid tests and number of tests performed).

The value of the dependent variable at a certain value of the independent variable (e.g. the number of positive Covid tests at a certain number of tests performed). [1]

1.2 Simple Linear Regression Model

1.2.1 Model significance

Considering a sample of variables as the following:

Length	X	Y
1	825	3,5
2	215	1
3	1070	4
4	550	2
5	480	1
6	920	3
7	1350	4,5
8	325	1,5
9	670	3
10	1215	5

Regression line equation: y=0.1181+0.0036x

Significance model with Anova

H0: b=0 H1:b<>0

	df	SS	MS	F
Regression	1,0000	16,6816	16,6816	72,3959
Residual	8,0000	1,8434	0,2304	
Total	9,0000	18,5250		

Degree of Freedom Regression -> It is always 1

Degree of Freedom Regression -> The number of observations minus 2.

Degree of Freedom Total -> The number of observations minus 1.

$$SR = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

Figure 1 SR Formula [2]

SR -> We calculate the summation of the y values calculate through the linear regression model minus the average of the y value itself to the power of two.

$$SE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

Figure 2 SE Formula [2]

SE -> We calculate the summation of the observed y value minus the y value from the linear regression model to the power of two.

MSR -> It is the result obtained by the division of SR and by the degrees of freedom from Regression.

MSE -> It is the result obtained by the division of SE and by the degrees of freedom from the Error.

F-test Statistics -> It is the result between the division of MSR by MSE.

Coefficient of Determination:

To determine the coefficient of determination it is necessary to firstly calculate the Sxy to the power of two, Sxx and Syy.

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

Figure 3 R Squared Formula [2]

Sxx-> Is the summation of each x value minus the average of the x values to the power of 2.

Syy-> Is the summation of each y value minus the average of the y values to the power of 2.

Sxy -> Is the summation of each x value minus the average of all x values multiplied by the result of subtraction of each y value with the average of all y values

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Figure 4 S formulas [2]

In our example this calculation looks like the following:

$$R^2 = \frac{21650409}{1297860 * 18{,}525}$$

R^2= 0.9004923770185123

With the value of R^2 we get valuable information about the behaviour of our sample. When the R^2 is close to 1 we can affirm that there is a degree of linearity between the y and x values. A value of 0 indicates that there is no relationship. [3]

Applying this to our example we get this graph which clearly shows that our points are not that incoherent with the line and therefore it supports the idea that when R^2 is closer to 1 there is a strong relationship between the two variables. Additionally, R^2 is relevant in predicting values since it can be used as a measure of the error. Higher R^2 values correspond to models with less error, which in turn produces predictions that are more precise. [4]

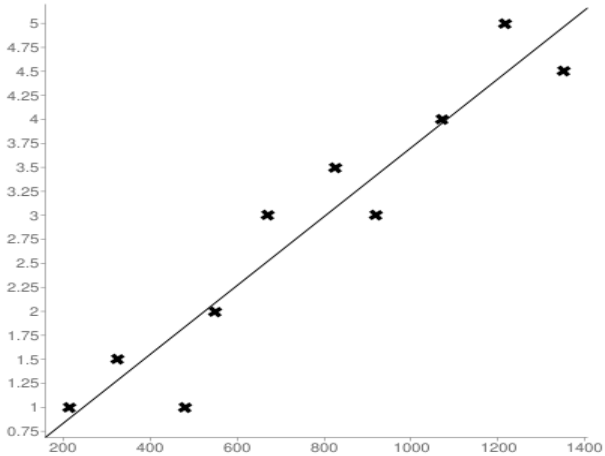


Figure 5 The model equation line and its points compared to it.

Other statistics:

R2adjusted = 0.8880

R = 0.9489

1.2.2 Hypothesis tests for model coefficients

Hypothesis tests for 5 significance level

H0: a=0.0 (b=0.0) H1: a<>0.0 (b<>0.0)

tobs(a)=0,3326 tobs(b)=8,5086

Decision:

Regarding the test for the a parameter h0 is not rejected

Regarding the test for the b parameter h0 is rejected

Let us break this information into pieces:

We are testing whether the parameter a and b being 0 or different to 0 can be supported by the data of the sample. For a parameter we need to calculate the tobs which stands for "t observed". To do this we will be using the formula which is presented in the image on the right. Whereas for the b parameter we need to calculate the formula beneath the one for the "a" parameter.

$$T_a = \frac{\hat{a} - a_0}{S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n\ -2}$$

Figure 6 Ta formula [2]

$$T_b = \frac{\hat{b} - b_0}{S/\sqrt{S_{xx}}} \sim t_{n\ -2}$$

Figure 7 Tb formula [2]

With that we need to formulate our decision. For this we will need to compare each tobs with the tStudent value for the significance value decided, which in our case is 5%.

T-Student Value for 5% significance value -> 2.3060 (which can be found in tables with tabulated values)

Since the tobs for the "a" parameter is less than the t-student value we can conclude that we cannot reject h0, whereas for the "b" parameter, since the tobs is more than 2.3060 we can reject the idea that we were trying to prove.

Hypothesis tests for 1 significance level

Hypothesis tests for regression coefficients

H0: a=0.0 (b=0.0) H1: a<>0.0 (b<>0.0)

tobs(a)=0,3326 tobs(b)=8,5086

Decision:

Regarding the test for the a parameter h0 is not rejected

Regarding the test for the b parameter h0 is rejected

For this example, we are trying to prove the same as the example before but this time with a confidence value of 1%. The steps to get to the tobs for both parameters are the same; however, since the confidence value is different, the T-Student Value which we will compare with both tobs will also be different. In this case the T-Student Value is 3.3554.

In a nutshell, since the tobs(a) is less than the T-Student Value we cannot reject, whilst because tobs(b) is more than the T-Student Value we can reject it.

With that we can conclude that we should reject the idea of h0 being 0 or different to 0 for the "b" Parameter and we cannot reject it for the "a" parameter.

1.2.3 Confidence intervals for prediction values

This represents how to do the elaboration of confidence intervals.

$$\left]\hat{y}_0 - t_c s\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}, \hat{y}_0 + t_c s\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\right[$$

Figure 8 Confidence intervals formula [2]

Where $\hat{y}_0$ represents the result of substituting the x variable in the linear model equation by each x value on the sample which is then subtracted or added with the result of the multiplication of Tc who plays for the T-student value which is calculated using the confidence value and s for the square root of the formula beneath the IC .

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Figure 9 s^2 formula [2]

Afterwards we need to do the squared root of 1/n where n represents the length of the sample in hands. X0-x is the value of the x being used minus the averege of all x values to the power of 2 divided by Sxx, which as said before it is the summation of each x value minus the average of the x values to the power of 2.

All these calculations for this example is defined below:

X values	95% confidence value intervals	90% confidence value intervals
825	]0.2525,1.5253 [Δ 0.6364	] 2.5588, 3.5923 [Δ 0.517066
215	] 3.954,4.4148[Δ 0.4605	] -0.0371, 1.8149 [Δ 0.9260
1070	]1.6838,2.4961 [Δ 0.4062	] 3.2841, 4.6243 [Δ 0.6701
550	]1.3945, 2.2835[Δ 0.4445	] 1.4990, 2.6809 [Δ 0.5910
480	]3.03422, 3.7987[Δ 0.3823	] 1.1922, 2.4858 [Δ 0.6468
920	]4.2880,5.6281[Δ 0.6700	] 2.8603, 3.9726 [Δ 0.5562
1350	]0.7330, 1.834[Δ 0.5503	]3.9831, 5.9330[Δ 0.9745
325	] 0.4826,1.8336 [Δ 0.8007	] 0.4826,2.0840 [Δ 0.8007
670	] 0.7330,2.8815[Δ 0.5503	] 1.9945, 5.2924[Δ 0.5257
1215	]3.9117, 5.0364[Δ 0.5624	] 3.6558, 5.2924[Δ 0.8183

2.Multilinear Regression

2.1 Overview of Multiple Linear Regression

Multilinear regression serves a purpose like simple linear regression, however instead of trying to make a model from one variable's relation to the other, it tries to predict a result using 2 or more independent variables.

2.2 Multiple Linear Regression Model

2.2.1 Model significance

Source	Degrees of Freedom	Sum of squares	Mean Square	F
Model	p	$\sum(\hat{y}_i-\overline{y})^2$	SSM/DFM	MSM/MSE
Error	$n - p - 1$	$\sum(y_i-\hat{y}_i)^2$	SSE/DFE	
Total	$n - 1$	$\sum(y_i-\overline{y})^2$	SST/DFT	

Figure 10 Model Significance [9]

USED VALUES FOR CALCULATION

n- number of sets of data used in the demonstration.

Y Matrix- a column matrix composed of the Y variable sets of data.

X Matrix- a matrix with rows equaling the amount of data samples collected and columns equaling the number of variables used.

$$\begin{bmatrix} 1 & x_{11} & x_{12} & \vdots & x_{1k} \\ 1 & x_{21} & x_{22} & \vdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \vdots & x_{nk} \end{bmatrix}$$

Figure 11 The Matrix used for the Calculations [10]

B Matrix - the matrix that contains the regression coefficients.

CALCULATING THE REGRESSION COEFFICENTS.

To calculate the regression coefficents we will first need to create an X Matrix that contains on the first column only the number one, and on the remaining columns all the Y values. After that we will need to transpose it. After that we will need to multiply in the following order,X transposed times X, and we will be given a square matrix with the size of the number of coefficents. After that we need to calculate the inverse of the result of that multiplication. Once that multiplication is finished we need to make a multiplication of the X Transposed matrix times the Y matrix. To finish the calculation we need multiply the inverse of the first multiplication to

the result of the second multiplication. That will give us a column matrix named B with the coefficients.

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T Y$$

Figure 12 The Design Matrix [10]

ANOVA

DEGREES OF FREEDOM

Regression Model Degrees of Freedom (DFr)-> is equal to the number of different variables used in the model.

Error Degrees of Freedom (DFe)-> is equal to the number of provided samples of data minus the number of variables minus one.

Total Degrees of Freedom (DFt) -> is equal to the sum of the regression model degrees of freedom and the error degrees of freedom.

SUM OF SQUARES

Regression Model sum of squares (SSr)- is equal to the matricial calculus of the B Matrix Transposed times the X Matrix transposed times the Y matrix which will give origin to a matrix with only one number . That number minus the the multiplication of the average of the y data to the power of two and n will give us the SSr Number.

$$SQ_R = \hat{\boldsymbol{\beta}}^T X^T Y - n\bar{y}^2$$

Figure 13 SQR formula [10]

Error sum of squares(SSe)- is equal to the matricial calculus of the Y Matrix times the Y matrix Transposed minus the multiplication of the B matrix transposed times the X matrix transposed times the Y matrix which will give origin to a matrix with a single number, that number being the SSe.

$$SQ_E = Y^T Y - \hat{\boldsymbol{\beta}}^T X^T Y -$$

Figure 14 SQE Formula [10]

Total sum of squares (SSt) – is equal to the sum of SSr plus SSe.

Mean Squares

Regression Mean Square(MQr) -is equal to the SSr divided by the DFr.

Error Mean Square(MQe) -is equal to the SSe divided by the DFe.

F-> is equal to the MQr divided by the MQe. Is also later used in a test to determine if the regression model is significant or not.

2.2.2. Hypothesis tests for model coefficients

$$H_0 : \beta_2 = 0 \quad \text{v.s.} \quad H_1 : \beta_2 \neq 0$$

Figure 15 The hypothesis tests being tested [10]

Hypothesis test for the third regression coefficent in a multi linear regression.

To make an hypothesis test about a coefficient we need to make the following calculation.

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

Figure 16 The test statistic formula [10]

The B is the coefficient that we want to calculate, the standard deviation to the power of 2 is equal to the MQe and the Cjj is obtained by going to the inverse of the X Transposed Matrix times X matrix and choosing the correspondent cell to the coefficient that you want to calculate. After performing this calculation. After that we need to get the proper t Student from the table using the desired significance divided by 2 and the Error Degrees of Freedom.

$$|t_0| > t_{0,975,[9]}$$

Figure 17 Rejection Criteria [10]

In this example the T is gotten from a significance level of 95% and 9 degrees of freedom.

If the the Absolute value of the t calculation is bigger than the t gotten from table, H0 will be rejcected,if not then the opposite is true.

1% significance.

```
Hypothesis tests for regression coefficients
H0: b0=0 (x1=0) (x2=0) H1: b0<>0 (x1<>0) (x2<>0)
 tobs(b0)=-4,4905 tobs(b1)=7,7196 tobs(b2)=1,5188
Decision:
Regarding the test for the b0 parameter h0 is rejected
Regarding the test for the b1 parameter h0 is rejected
Regarding the test for the b2 parameter h0 is rejected
```

The application showed us this which means that both the hypothesis were rejected with a significance level of 1%

5% significance

```
Hypothesis tests for regression coefficients
H0: b0=0 (x1=0) (x2=0) H1: b0<>0 (x1<>0) (x2<>0)
 tobs(b0)=-4,4905 tobs(b1)=7,7196 tobs(b2)=1,5188
Decision:
Regarding the test for the b0 parameter h0 is rejected
Regarding the test for the b1 parameter h0 is rejected
Regarding the test for the b2 parameter h0 is rejected
```

The application showed us the same result which means that in this case the difference in significance did not make a difference.

2.2.3 Confidence intervals for prediction values

To calculate the confidence interval for a prediction value we need to first calculate the value using the regression model line.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Figure 18 The line equation [10]

After doing that to calculate the confidence interval we need to calculate the following equation:

$$t_{1-\alpha/2[n-(k+1)]} \sqrt{\hat{\sigma}^2 x_0^T C x_0}$$

Figure 19 Delta [10]

First we need to get the t of the significance level we want to choose, after that we can multiply it to the value of the anova error degrees of freedom . The remaining equation is obtained through the square root of MQE times a matricial equation: the matricial equation is a row matrix with a 1 in the first cell and the x values used to make the prediction In the remaining ones. We take that row matrix and multiply it to the X Transposed Matrix times X Matrix. After that we will get another row matrix which will be multiplied to a column matrix with the same values as the first matrix. That will create a matrix with a single number that will be used in the calculation of the confidence interval.

After calculating the equation we simply have to create an interval with the left side having the predicted value minus the equation, while on the other side we have to add the value instead of subtracting it.

$$\left]\hat{\mu}_{Y_0} - t_{1-\alpha/2[n-(k+1)]} \sqrt{\hat{\sigma}^2 x_0^T C x_0} \;,\; \hat{\mu}_{Y_0} + t_{1-\alpha/2[n-(k+1)]} \sqrt{\hat{\sigma}^2 x_0^T C x_0}\right[$$

Figure 20 The Confidence Intervals Formula [10]

95% confidence intervals

The only thing that changes with the significance numbers is the t obtained from the table.

14,32	14,31-14,33
9,30	9,28-9,31
9,66	9,64-9,68

Example of 95% intervals in our application

90% confidence intervals.

14,32	14,29-14,35
9,30	9,26-9,33
9,66	9,62-9,70

Same example as above but this time with 90% intervals

Annex B MDISC

1 Sorting tests results by client name or TIN

1.1 Introduction

Sorting is a process of rearranging a list of elements to the correct order since handling the elements in a certain order is way more efficient than handling randomize elements [5]. Our problem at hands involves exactly that. The Clinical Chemistry Technologist wants to see the test details of the all the tests associated to a target client. For this, he firstly needs to select the target client to see all his test results. Having said that, it is natural to say that it can be difficult to select a client in a list containing thousands of entries without any organization criteria. Therefore, ManyLabs has decided that the clients can be organized in two criteria: Tax Identification Number (Tin) and his name. Additionally, it was requested that the application featured two sorting algorithms.

Selection Sort:

ALGORITHM 1: ALGORITHM TO SORT THE CLIENTS (SELECTION SORT)

```
procedure selection sort
  list  : array of clients
  n     : size of list

  for i = 1 to n - 1
  /* set minimum as the current element*/
    min = i

    /* check the element to be minimum */

    for j = i+1 to n
      if list[j] < list[min] then
        min = j;
      end if
    end for

    /* swap the minimum element with the current element*/
    if indexMin !=  i  then
      swap list[min] and list[i]
    end if
  end for

end procedure
```

Selection Sort **gets the element in the 1st position** of an unsorted array and scans through subsequent elements to **find the smallest element**. Once the smallest element is found, it is swapped with the element in the 1st position. The algorithm then moves on to the element in the

2nd position and scans through subsequent elements to find the index of the 2nd smallest element. Once found, the second smallest element is swapped with the element in the 2nd position. [6]

This process goes on until we reach the n-1st element of the array, which puts the n-1st smallest element in the n-1st position. The last element automatically falls in place since there are no more elements to be compared to and thus completing the algorithm.

Reverse order: We find **the largest element instead of the smallest element to sort the array in descending order.**

Insertion Sort:

ALGORITHM 2: ALGORITHM TO SORT THE CLIENTS (INSERTION SORT)

```
procedure insertionSort( A : array of items )
  int holePosition
  int valueToInsert

  for i = 1 to length(A) inclusive do:

    /* select value to be inserted */
    valueToInsert = A[i]
    holePosition = i

    /*locate hole position for the element to be inserted */

    while holePosition > 0 and A[holePosition-1] > valueToInsert do:
      A[holePosition] = A[holePosition-1]
      holePosition = holePosition -1
    end while

    /* insert the number at hole position */
    A[holePosition] = valueToInsert

  end for

end procedure
```

An insertion sort compares values in turn, **starting with the second value** in the list. If this value **is greater than the value to the left of it**, no changes are made. Otherwise, this value is repeatedly moved left until it meets a value that is less than it. The sort process then starts again with the next value. This continues until the end of the list is reached. This sorting method is often associated to the way you sort playing cards in your hands. The array is virtually split into a sorted and an unsorted part. Values from the unsorted part are picked and placed at the correct position in the sorted part.[7]

1.2Runtime tests for inputs of varying sizes

n	Selection Sort unsorted	Insertion Sort unsorted
...	...	...
10000	0,29 s	0,30 s
20000	1,09 s	1,36 s
30000	2,46 s	3,71 s
40000	4,54 s	7,71 s
50000	7,13 s	14,23 s
60000	8,5891 s	26,60 s
…	…	…
100000	29,4123931s	45.009,7 s

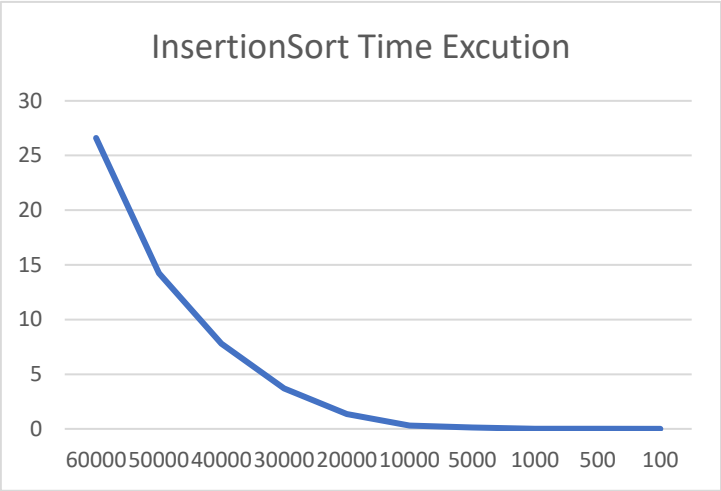


Figure 22 InsertionSort Asymptotic Behaviour

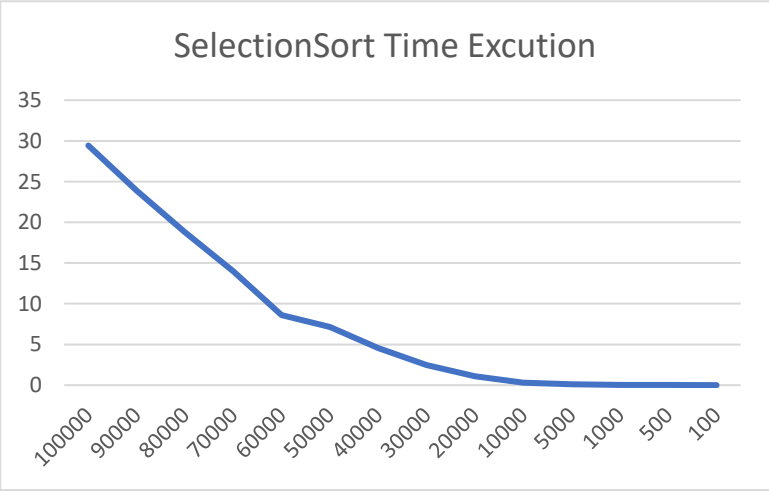


Figure 21 SelectionSort Asymptotic Behaviour

1.3 Worst-case time complexity analysis

Selection Sort

Theoretically, the search for the smallest element should always take the same amount of time, regardless of the initial situation. And the swap operations should only be slightly more for elements sorted in descending order (for elements sorted in descending order, every element would have to be swapped; for unsorted elements, almost every element would have to be swapped).

With elements sorted in descending thorder, we have – as expected – as many comparison operations as with unsorted elements – that is, $n \times (n-1) / 2$.
With unsorted elements, we have – as assumed – almost as many swap operations as elements: for example, with 4,096 unsorted elements, there are 4,084 swap operations. These numbers change randomly from test to test.
However, with elements sorted in descending order, we only have half as many swap operations as elements! This is because, when swapping, we not only put the smallest element in the right place, but also the respective swapping partner.
With eight elements, for example, we have four swap operations. In the first four iterations, we have one each and in the iterations five to eight, none (nevertheless the algorithm continues to run until the end):

The reason why Selection Sort is so much slower with elements sorted in descending order can be found in the number of local variable assignments (minPos and min) when searching for the smallest element. While with 8,192 unsorted elements, we have 69,378 of these assignments, with elements sorted in descending order, there are 16,785,407 such assignments – that's 242 times as many!

Insertion Sort

In the worst case, the elements are sorted completely descending at the beginning. In each step, all elements of the sorted sub-array must, therefore, be shifted to the right so that the element to be sorted – which is smaller than all elements already sorted in each step – can be placed at the very beginning.

In the following diagram, this is demonstrated by the fact that the arrows always point to the far left:

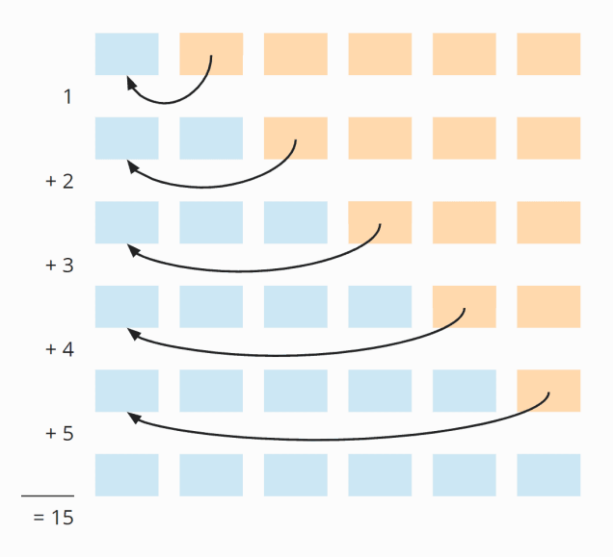


Figure 23: Diagram showing worst case scenario [8]

The term from the average case, therefore, changes in that the second dividing by two is omitted:

$6 \times 5 \times ½$

Or:

$n \times (n - 1) \times ½$

When we multiply this out, we get:

$½ n^2 - ½ n$

Even if we have only half as many operations as in the average case, nothing changes in terms of time complexity – the term still contains n^2, and therefore follows:

The worst-case time complexity of Insertion Sort is: $O(n^2)$

2. Evaluation of the effectiveness of the company's response

2.1 Introduction

To evaluate the performance of the company, Many Labs wanted to have a functionality where it would be simple so check it. For this, we check the sub-intervals in which there were more samples waiting for the result than samples already with a result attributed. This functionality is coordinated by the laboratory coordinator, which wants an overview of all the tests performed by Many Labs. Accordingly for the performance analysis it is needed to use at least two algorithms (a benchmark already provided by Many Labs and a brute force algorithm that was developed by our team).

ALGORITHM 1: BRUTE FORCE ALGORITHM

```
procedure brute force
  nums  : array containing the difference between the tests registered and tests validate in time windows of 30 minutes
  n  : size of list
  maximumSubArray : minimum value possible in the array
  start : 0
  end : 0
  count : 0

  for left = 0 to left < n
  /* set runningWindowSum as 0*/
    runningWindowSum = 0

    /* check if runningWindowSum is more than maximumSubArraySum*/

    for right = left to right < n
      runningWindowSum = runningWindowSum + nums[right]
      if runningWindowSum > maximumSubArraySum then
        start = left
        end = right
      end if
    end for
  end for

  new : array of integers initialized in which its length is the result of end value – start value + 1

  while start<=end then
    new[count] = nums [start]
    start = start + 1
    count = count + 1
  end while

  return new

end procedure
```

This algorithm produces an array of integers containing the maximum subsequence of the array introduced (nums). This is performed by firstly assigning the initial values to the variables used in the algorithm. Afterwards we go through all left variables which is incremented until it

is equal to n.< and assign the runningWindowSum variable to 0. Subsequently we assign right as being left and go through right until it is equal to n. With this we now increment the runningWindowSum variable with the element in the position of the value of right of the nums array. Then we check whether runningWindowSum is more than maximumSubArraySum and if it is true, we assign start as being left and end as being right, otherwise we go back to the beginning of the loop and repeat the same steps. After having gone through all the loops we create a new variable (new) in which its length is the result of end value – start value + 1. And while start is less or equal to end, we assign the value that is in the start position of the nums array to the position of the variable count in the new array, we then increment start by one and count by one.

Aftwards we return the array which was the result of this procedure.

2.2 Runtime tests for inputs of varying sizes

n	Benchmark	Brute force
100	0,0007	0,0001466
200	**0,0006809**	**0,0006805**
300	0,0006976	0,0008185
400	**0,0007415**	0,0013281
500	0,0007223	0,0017584
600	**0,0007556**	0,0019557
700	**0,0007515**	0,0020655
800	**0,0007546**	0,0024427
900	0,0008389	0,0034654
1000	0,0011943	0,0041385

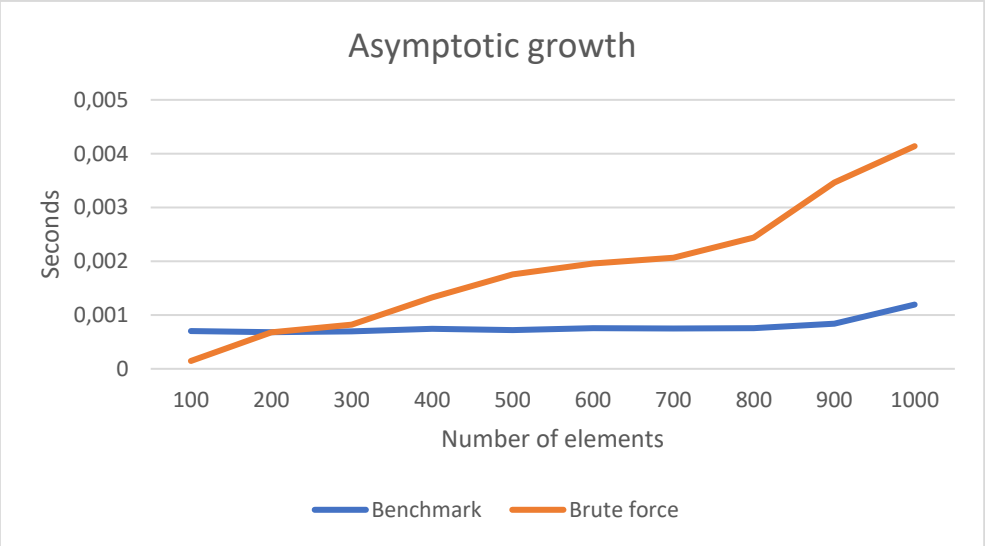


Figure 24 Asymptotic growth of both algorithm

2.3 Worst-case time complexity analysis

ALGORITHM 1: BRUTE FORCE ALGORITHM

```
procedure brute force
  nums  : array containing the difference between the tests registered and tests validate in time windows of 30 minutes
  n  : size of list
  maximumSubArray : minimum value possible in the array
  start : 0
  end : 0
  count : 0

  for left = 0 to left < n
  /* set runningWindowSum as 0*/
    runningWindowSum = 0

    /* check if runningWindowSum is more than maximumSubArraySum*/

    for right = left to right < n
      runningWindowSum = runningWindowSum + nums[right]
      if runningWindowSum > maximumSubArraySum then
        start = left
        end = right
      end if
    end for
  end for

  new : array of integers initialized in which its length is the result of end value – start value + 1

  while start<=end then
    new[count] = nums [start]
    start = start + 1
    count = count + 1
  end while

  return new

end procedure
```

Presenting our pseudocode again we can now calculate the big O notation for our algorithm. For this we will need to do the following operations:

1-Break your algorithm/function into individual operations

2-Calculate the Big O of each operation

3-Add up the Big O of each operation together

4-Remove the constants

5-Find the highest order term — this will be what we consider the Big O of our algorithm/function

Let us firstly break our algorithm into individual operations:

Operation	Big O for each operation
Look up for nums	O(1)
Assign n to the size of nums	O(1)
Assign maximumSubArray with the minimum number possible	O(1)
Assign start to 0	O(1)
Assign end to 0	O(1)
Assign count to 0	O(1)
Loop through left and right until it reaches n-1	O((n-1)(n-1)) =O(n^2)
Assign runningWindowSum to 0	O(n-1)
Increment runningWindowSum by the value in the position right of the array nums	O(n-1)(n-1)= O(n^2)
Check if runningWindowSum more than maximumSubArraySum	O(n-1)(n-1)= O(n^2)
Assign the new array with the length of end value – start value + 1	O(1)
Assign the value in the position start of the nums array to the value in the position count of the new array	O(n)
Increment start by 1	O(n)
Increment end by 1	O(n)
Return new	O(1)

We can now add up each big O operation and we get :

O(1+1+1+1+1+1+(n-1)(n-1)+n-1+ (n-1)(n-1)+ (n-1)(n-1)+1+n+n+1) = O(3n^2-3n+10)

Now with this we can extract the highest order term which is 3n^2.

Now we can generalize O(3n^2) to O(n^2), since O(kn^i) is =(n^i) and therefore we can conclude that this algorithm has a big O notation of O(n^2).

References

[1] Adapted from Milne Publishing. [Online] (2021). Available: https://milnepublishing.geneseo.edu/natural-resources-biometrics/chapter/chapter-7-correlation-and-simple-linear-regression/

[2] From Isep Moodle MATCP. [Online] (2021). Available: https://moodle.isep.ipp.pt/pluginfile.php/129148/mod_resource/content/5/Teo%CC%81rica%207.pdf

[3] Adapted form PennState Ebelry College of Science STAT 462. [Online] Available: https://online.stat.psu.edu/stat462/node/98/

[4] Adapted from StatisticsByJim. [Online] (2018). Available: https://statisticsbyjim.com/regression/interpret-r-squared-regression/

[5] Adapted from StudyLib. [Online] (2018). Available: https://studylib.net/doc/25386078/analysis-and-comparative-of-sorting-algorithms

[6] Adapted from Wikipedia SelectionSort. [Online] (2021). Available: https://www.wikiwand.com/en/Selection_sort

[7] Adapted from BBC Common Algorithms. [Online] [2021] Available: https://www.bbc.co.uk/bitesize/guides/zjdkw6f/revision/6

[8] Adapted from Happy Coders. [Online] [2020] Available: https://www.happycoders.eu/algorithms/insertion-sort/#Worst-Case_Time_Complexity

[9] Adapted from Yale University. [Online] [1997] Available: http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

[10] Adapted from ISEP Moodle MATCP. [Online] [2021] Available: https://moodle.isep.ipp.pt/pluginfile.php/129989/mod_resource/content/5/Teo%CC%81rica%208.pdf