DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Development of a Prediction App, Which Predicts If a Person Intends to Do Sport Soon.

Alexander Ruhl

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Development of a Prediction App, Which Predicts If a Person Intends to Do Sport Soon.

Entwicklung einer Prädiktions-App, welche vorhersagt, ob eine Person vorhat, demnächst Sport zu machen.

Author:	Alexander Ruhl
Supervisor:	Prof. Stephan Jonas
Advisor:	Jonas Schulte-Coerne
Submission Date:	16.3.2020

I confirm that this bachelor's thesis in informatics is my own work and I have documented all sources and material used.

Munich, 16.3.2020 Alexander Ruhl

Abstract

The goal of this bachelor thesis is to develop a concept for an app that can predict whether someone intends to do sport soon. This concept was created by evaluating the data that was collected in a study carried out for this purpose. One part of the bachelor's thesis was to develop an app that collects data for the study. Furthermore a list of possible indicators for the prediction of sport was collected. The study showed that sport can be predicted using the following indicators: the **plan to do sport today**, the **day of the week**, the **number of sports units per week**, the **sleep**, the **satisfaction with meals**, especially with the first meal and the **typical time of a day at which sport is done**. With these indicators, the developed concept can predict with an accuracy of 84% whether sport will take place.

Kurzfassung

Das Ziel dieser Bachelorarbeit ist es, ein Konzept für eine App zu entwickeln, welche vorhersagen kann, ob jemand vorhat, demnächst Sport zu machen. Dieses Konzept entstand durch die Auswertung der Daten, welche in einer hierfür durchgeführten Studie gesammelt wurden. Teil der Bachelorarbeit war es, eine App zu entwickeln, welche Daten in der Studie sammelt. Des Weiteren wurde für das Sammeln der Daten eine Liste an möglichen Indikatoren für die Vorhersage von Sport zusammengetragen. Die Studie hat ergeben, dass mit folgenden Indikatoren Sport vorhergesagt werden kann: das **Vorhaben heute Sport zu machen**, der **Wochentag**, die **Anzahl an Sporteinheiten pro Woche**, der **Schlaf**, die **Zufriedenheit mit Mahlzeiten**, insbesondere mit der ersten Mahlzeit und die **Tageszeit**. Mit diesen Indikatoren kann das entwickelte Konzept zu einer Genauigkeit von 84% vorhersagen, ob Sport stattfinden wird.

Contents

Abstract iii

Kurzfassung iv

1 Introduction 1

2 Study 2
2.1 Indicators . . . 2
2.2 Procedure of the Study . . . 3
2.3 Overview of the Collected Data . . . 4

3 Implementation 5
3.1 General . . . 5
3.2 Design . . . 7
3.3 Questionnaire . . . 8
3.4 Adding Data . . . 8
3.5 Weather . . . 8
3.6 Upload the Data . . . 9
3.7 Encryption and Decryption . . . 9

4 Evaluation 10
4.1 Age . . . 10
4.2 Regularity of Sport . . . 10
4.3 Number of Sports . . . 10
4.4 Number of Sport Units Per Week . . . 11
4.5 Fixed Times or Usual Times . . . 11
4.6 Day of the Week . . . 11
4.7 Daily Structure . . . 12
4.8 Daytime . . . 12
4.9 Plan to Do Sport Today . . . 12
4.10 Unplanned Sport . . . 12
4.11 Sleep . . . 13
4.12 Weather . . . 13
4.13 Satisfaction with Meals . . . 14
4.13.1 Satisfaction with the First Meal . . . 14
4.14 Gender . . . 14
4.15 Further Possible Indicators . . . 15

5 Prediction **16**
5.1 Whether-Prediction 16
5.1.1 Plan to Do Sport 16
5.1.2 Day of the Week 17
5.1.3 Number of Sport Units Per Week 17
5.1.4 Sleep 17
5.1.5 Satisfaction with Meals 17
5.2 When-Prediction 18
5.3 Evaluation of the Prediction Concept 18

6 Outlook **20**

Bibliography **21**

1 Introduction

More than half of the population in Germany does sports at least once a week [1]. Furthermore nutrition is closely linked to sport. Therefore the question arises to what extent sport influences nutrition.
The sports and health scientists at the Technical University of Munich are researching whether sporting activity has an influence on the choice of your own meals and whether you eat more appropriately if you commit to your diet before exercising. So far, the results of these scientists are that a differential relationship between habitual physical activity and dietary intakes, whereby moderate but not necessarily highest physical activity levels are associated with reduced added sugar and increased nutrient-dense food consumption [2]. Furthermore, their research shows that in normal-weight women, calorie restriction and exercise result in weight loss, loss of fat mass and characteristic metabolic adaptations indicative of energy conservation. However, weight loss was considerably less than expected from advanced prediction models, and contrary to previous literature, fat-free mass was almost completely preserved in their group of normal-weight women[3]. So far, they have only carried out their experiments under very controlled laboratory conditions, which severely limits the number of test subjects and may influence the results. The aim of the bachelor thesis is to develop concept for a sporting activity prediction app. This would be a building block that would allow the sports and health scientists to collect relevant research data in the everyday life of their subjects.
The aim of the thesis is to predict sport or to find out to what extent sport can be predicted. At the time of the bachelor's thesis, there is no known research to predict sport. Therefore, many assumptions are made about the indicators for doing sport. These indicators are then tested on the basis of a short-term study. Part of the bachelor thesis is to develop an app with which the data can be collected. This data is then evaluated and an algorithm is developed based on the indicators that predicts sport.
In the course of this thesis, the planning and course of the study are described. The choice of indicators is explained and further possible indicators are discussed. Then it shows how the app for data collection was implemented. After this the data collected from the study gets evaluated and interpreted. Then, based on the interpreted data, an concept is developed to predict sport.

2 Study

2.1 Indicators

The aim of the work is to develop an app that predicts sport. The first question is how to predict sport. For this purpose, indicators that could be relevant for the prediction were collected on the basis of personal experience and a survey of the test subjects before the experiment. On this basis, the following **indicators** were selected to validate their relevance to sport prediction in the short-term study. The following indicators have been selected:

- **Age:** This indicator shall determine possible age differences in sports behavior.
- **Regularity of Sport:** The regularity of sport is a point value with which the test person evaluates the regularity himself. The purpose of this indicator is to check how often an athlete observes the times he set for sport.
- **Number of Sports:** This could affect the frequency of sport and some sports may be done more often than others.
- **Number of Sport Units Per Week:** This should create a basis for judging how often a person does sport.
- **Fixed times:** If a person trains at set times a week due to club sports or similar, this can also serve as an indicator.
- **Daily Structure:** This indicator is used to check to which extent sport is done to set time windows, this indicator is aimed at the sports in which the athlete freely allocates the time he does sport (for example strength training).
- **Daytime:** This indicator is used to help narrow down the exact time of a day at which sport is done.
- **Day of the Week:** This indicator is used to help narrow down the day on which sport is done.
- **Weather :** The weather has been tracked to check if the weather has an impact on sporting activity. It is checked whether less sport is done on days with rain in comparison to days without rain.
- **Sleep:** Little sleep can lead to less energy, which can lead to a person not doing any sport.

- **Plan to Do Sport Today:** Doing sport probably follows from the intent of doing sport.
- **Satisfaction with Meals:** Dissatisfaction with meals could make a person feel bad, which could affect doing sport.
- **Satisfaction with the First Meal:** The first meal, as the lead-in of the day, could have a larger impact on sport.

The following indicators have not been recorded, the reasons for this are explained here.

- **Season:** According to the survey of the test persons, there are people who exercise more in summer, mostly outdoor sports such as jogging. On the other hand, there is also the group of people who find it too warm in the summer and who prefer to go to the gym in the cold season. Due to the short duration of the study, the season could not be taken into account.
- **Pedometer** With this indicator the activity level of a person could be determined and depending on this, further information could possibly be obtained for a prognosis. Due to the high expenditure of time for the implementation of pedometer for Android, this indicator was not validated in the study. Android offers a step detector sensor and a step counter sensor for this, however, the following problems arise. The app must always remain open in the background and the counter resets when the app is closed. This means that it also resets when restarting [4]. In addition, this would result in high battery consumption, which would not be user-friendly. An API for Google Fit offers an alternative to this, but there were not enough Google Fit wristbands available for the study.

2.2 Procedure of the Study

The task of the subjects was to enter data into the app developed for this purpose every day over a period of four weeks. Seven people participated in the study, but two people broke off (N = 5). When starting the app for the first time, the following personal information should be provided:

- age
- regularity of sport (rating from 1, very irregular, to 7, very regularly)
- the total number of sports
- the average number of sport sessions per week
- postal code of the place of residence
- fixed times of doing sport (weekday and time)

- how a day of the subject is structured (rating from 1, very unstructured, to 7, very structured)
- usual time of doing sport (morning, noon or evening)

In addition to the one time questionnaire, the following data was collected daily:

- in the morning after getting up:
  - rating of sleep
  - plans to do sports today (yes, maybe, no)
- after a sport session, start and end of each sport session and whether it was planned
- after a meal, satisfaction with the meal (rating from 1 to 5)
- weather was also recorded daily in the background

2.3 Overview of the Collected Data

In the study with n = 5 subjects, the following amount of data was collected over a period of 29 days (145 days in total):

- 63 sport units including whether they were planned
- start and end time of each sport unit
- 251 meals on 131 days
- weather of 111 days
- rating of sleep on 133 days
- plan to do sport on 133 days
- all answers to the questionnaire of each subject

3 Implementation

The following chapter deals with the implementation of the app, which has been programmed to collect the data. The app was designed for Android. Accordingly, Android Studio (version 3.5.3.) was used as the programming environment. The programming language used is Java. The minimum requirement for the app is Android SDK version 15, the target version is Android SDK 29. The most important aspects of programming are discussed below.

3.1 General

When you start the app you get to the start screen. This is used to check whether the questionnaire has already been answered. If this is not the case, you will be forwarded to the questionnaire. After answering the questions you will be forwarded to the main menu. If the questionnaire has already been answered, you will be directly forwarded from the start screen to the main menu. From the main menu (see figure 3.2.a) you can access

- the `MorningActivity` to answer the morning questions,
- the `MealActivity` to add a meal,
- the `SportActivity` to enter a sport unit and
- via the question mark symbol in the `HelpAndUploadActivity` to either read how to use the app or to upload the data.

See Figure 3.1 for a visual representation.

The app has been divided into three packages for a better overview when programming, to separate functionalities and to subsequently change or add functions more easily (see figure 3.1). The first package `AskCatalog` contains the activities of the questionnaire. The second package `MainMenu` contains the activities that can be accessed through the main menu, including the main menu itself (see figure 3.2.a). The last package is the root package. This package contains the start activity and a separate Java program `Decrypt`. At the end of the study, after all data has been received by email, this Java program is then used to decrypt the collected data.

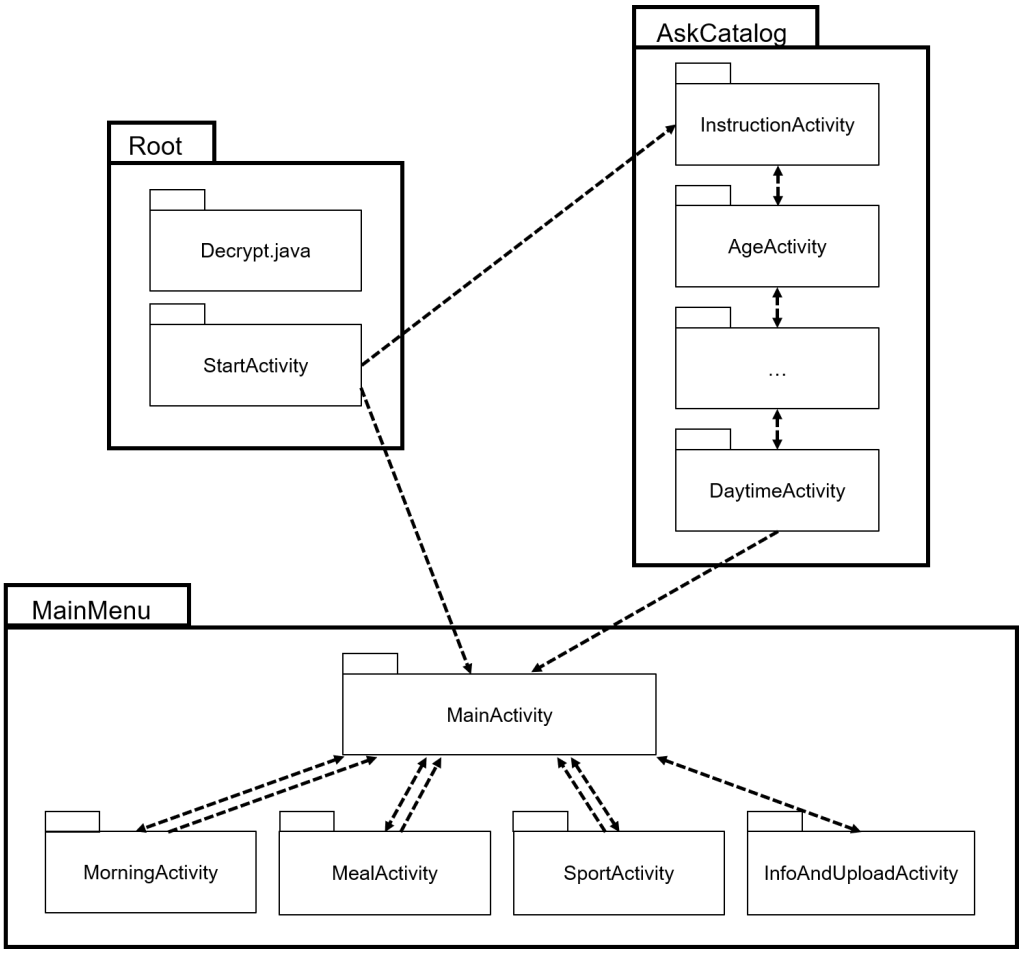
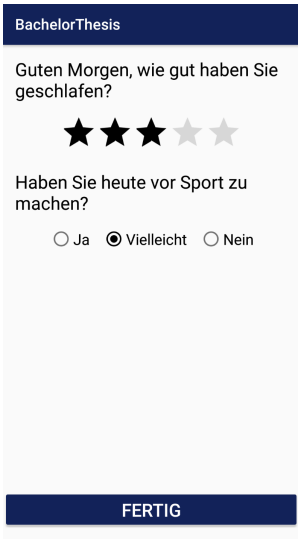


Figure 3.1: The package diagram shows the interaction of the individual activities and the affiliation to the packages. Each activity represents its own screen. If an arrow points in both directions, you can return to the previous screen with the back button on the smartphone.

(a) Main Menu



(b) Screen for entering data every Morning. Rating of sleep at topside and the intention of doing sport today below.

(c) Example widget of entering a date

Figure 3.2: Three example screens.

3.2 Design

The design was kept simple because the focus was primarily on the experiment. The main colors chosen were blue with a white background based on the colors of the Technical University of Munich (see figure 3.2.a and 3.2.b). All widgets were displayed in black. Here too, the focus was on simplicity, so that the test participants were not distracted. The high contrast made all texts easy to read and all buttons clearly visible (see figure 3.2.c). Texts were displayed in German within the app because, all test participants are native German speakers.
Furthermore, the questions are displayed individually on the questionnaire. So that the person only thinks about a single question at a time and do not have to scroll. Adding additional questions or changing the order is also made easier by this design.
In addition, the data is uploaded by clicking on the question mark. The question mark serves as a help symbol, which leads to an activity on which you can find an explanation of how to use the app. Since the upload of the data is a one-time action, this function was set up in this activity so that an upload button in the main menu does not distract unnecessarily. Furthermore there is no "back" button on the `HelpAndUploadActivity` so the layout looks clean. Returning to the main menu is done with the back button on the smartphone.

3.3 Questionnaire

When starting the app you get to the start screen. With a "next" button you get to the next screen. It is now checked in the background whether a CSV (Comma-separated values, file format to display tables) file has been created. If this is not the case, the questionnaire has not yet been completed successfully and the questionnaire is accessed via the start screen. Each question is displayed individually on a screen. Before going to the next question, it is checked whether the entry was correct. If this is the case, the next question is asked. The answers to the questions are stored in an object. After all questions have been answered the CSV file is created. After the CSV file has been created, you always get directly to the main menu via the start screen.

3.4 Adding Data

After the user of the app has entered data, it becomes added in the CSV file. A pop-up window will appear to visually signal to the user that the data has been added. In the CSV file a new line with the relevant information is added for each entry. When adding the morning questions, the *current date*, the *current time*, the *sleep rating*, the *training intent* and the *weather forecast for the current day* are saved. When adding a meal, the *entered date*, the *current time* and the *meal rating* are saved. Furthermore when a sport unit is added, the *entered date*, the *start and end time* and *whether it was planned* are saved.
The first line of the CSV file contains the name of the question and the answer. When evaluating the data, this line was initially removed so that the rest is clearly displayed in a spreadsheet program. The remaining lines are structured as follows. The date is in the first column. In the second column, the data entered then start in the same order as described above. In addition, not only the value was saved but also a key word about its meaning. So it was possible to search for the relevant key word when evaluating the data and accordingly only this data was displayed. For example, when evaluating meals, the keyword was "Essen" (German for food). When adding an entry, a new line was added to the end of the CSV file.

3.5 Weather

If the user clicks the button to answer the daily questionnaire in the morning, it is checked whether there is an internet connection. If this is not the case, a popup window appears asking the user to connect the smartphone to the internet. Only when the device is connected to the internet the user will be forwarded to the next screen for entering the data. This is done to ensure that the weather can be requested. Furthermore the query of the weather runs in the background to improve the user experience.
For better usability, the weather was not determined by GPS, but by postcode of the place of residence. Since this was a voluntary study, the focus was on user-friendliness in order to make the test subjects as little trouble as possible. If the weather was determined by GPS, the GPS on the smartphone must be switched on every time the app wants to determine the

weather and in addition more battery would be consumed. In addition, the query of the weather by postcode reduces the risk of privacy violations.
Furthermore it can happen that there is a connection to a router, but the router is not connected to the Internet. In this case, it is incorrectly assumed that there is an internet connection. This is a calculated risk, but happened few times in the course of the study. Nevertheless if this happens, the weather for the day is not saved.
To keep track of the weather, the "5 day weather forecast" API from "OpenWeatherMap" was used. Furthermore internet permission has to be granted to the app.
The weather is transmitted using a JSON file which contains the weather for the next five days in three hour sections. The respective weather of the current day is filtered out accordingly. After conversion into strings, these are then added to the CSV file.

3.6 Upload the Data

In the main menu, the question mark symbol leads to the `HelpAndUploadActivity`. On this activity there is a text explaining how to use the app. In addition, there is a button with which the CSV file can be sent via email. By clicking on the button, a window opens in which the file can be sent with an email program of the choice. After the email has been sent you will be directed back to the `HelpAndUploadActivity`.
The file provider has only access to the encrypted file. Additionally no other apps have access to this file to increase data protection. When clicking on "Upload file", the CSV file gets encrypted and then the encrypted file is sent.

3.7 Encryption and Decryption

After the four weeks, the data was encrypted and sent via email manually. After receiving all the data, they were mixed again so that the data can no longer be assigned to any person. In addition, no names were saved and the subjects were randomly numbered. The AES (Advanced Encryption Standard) algorithm was used to encrypt the data. The key is saved as a string with a length of 16 characters. The key is a randomly generated sequence of lower and upper case letters, numbers and special characters. The key is stored hard-coded in the app and applies to all test subjects. The data is decrypted with its own Java program regardless of the app. In order to decrypt a file, the input path (code line 20) to the file to be decrypted and output path (code line 40) for the encrypted file has to be changed in the Java program. Then the program can be executed.

4 Evaluation

The starting point for the bachelor's thesis is that there is no known research on sport prediction. On this basis, a list of possible indicators was drawn up. An evaluation of the data showed the following.

4.1 Age

Initially a participant who is 39 years old participated. However, this person canceled the study. The age of the other subjects are between 20 and 23. Due to the fact that they are in the same group of age, no valid differences in athletic behavior can be determined. This means that no information has been obtained for the prediction of sports for this indicator.

4.2 Regularity of Sport

In the evaluation of the regularity, there were no connections found between the prediction of sport and the regularity. This indicator was used to check whether people who claim to exercise more regularly exercise more often on the same days. And it was checked whether there is a connection between the actual sport per week value and the self-declared average sport per week value in relation to regularity. The problem of this indicator is that there is no distinction between phases in which regular sport is done and phases in which this happens more irregularly. According to a test participant, sport was done more regular in the month before the study and therefore a high value for regularity was entered. In the month of the study, sport was no longer done as regularly.

4.3 Number of Sports

The surveys include one person with four kinds of sport (volleyball, jogging, swimming and strength training), one person with two kinds of sport (yoga and strength training) and three persons with one kind of sport (strength training). Due to the small number of test participants, the differences between the three groups are not significant. Nevertheless, there is a tendency to see that people who do several sports do more sport per week than people who do only one sport. Individuals with only one sport exercised an average of 10 times during the study period, the person with two sports 18 times and the person with four sports 15 times. Since all people did strength training, it could be that the people with several sports did the other sports in addition. Since there was no distinction between the sports, no further

statements can be made. For this reason a distinction should be made between the sports for further studies.

4.4 Number of Sport Units Per Week

The test participants indicated that they exercise on average 3.6 times a week. The actual value of the sport units per week is 2.97% with the standard deviation of 0.59. As a result he average deviation from their own information is -0.628. This means that the value can be expected to provide a guideline for the actual number of sports per week.
A survey of the participants showed that this value was set as high as it was planned. However some of the participants sometimes did not do sport like planned. Furthermore, it can be seen that this value was set too high for one person and was not achieved in any week. In contrast, it also happened that one person exceeded this value every week.

4.5 Fixed Times or Usual Times

After a survey, the participants replied that none of them takes part in a club or comparable sport and no one had fixed times, so this indicator could not be checked. Nevertheless, three people set fixed times because they understood the question as if the fixed times were their usual times. Person 1 exercised five times in the specified period and ten times outside. Person 3 exercised six out of eight times in the given period and person 5 never. It seems as if the usual times have a low informative value. Perhaps these were the times that the test participants had set themselves at the beginning of the year, but could not turn over. On the other hand, it could be that if the test group were larger, most people like person 3, who complies with this to 75%, would have followed these times. However, an unfavorable period was also chosen here due to the turn of the year and and possible new year resolutions. Another possibility for a wrong result could be the small number of test participants distorting the result.

4.6 Day of the Week

It was evaluated for each person for each day of the week how often they do sport on that day. The evaluation has shown that most of the test persons exercise more often on some days than on others. It is assumed that some athletes do not have time for sport on certain days or use these days as rest days. For example, person 1 did not do sport on any Monday, but every Thursday. Furthermore, it was observed that some subjects did sport in certain periods on certain days. For example, person 5 always started doing sports between 6:30 p.m. and 7 p.m. on Wednesdays. This knowledge can help to determine a period in which sport is likely to take place. Since the study was only held over a period of four weeks, the data can still be very inaccurate, nevertheless it can be assumed that the day of the week plays a major role in the prediction of sport.

4.7 Daily Structure

There were no connections found between daily structure and sport prediction. This is because one person rated it with seven out of seven and the rest rated it with five out of seven. The question arises whether the person who rated their daily structure with 7 out of 7 did the most sport at the same time. Even if there was a tendency towards this in the person, this tendency was also evident in two other people with a rating of 5/7. Thus the test participants do not differ enough to determine differences.

4.8 Daytime

Each participant exercised on average 71.4% of the time at the preferred time of day. This percentage occurrences because one person only exercised 2 times out of 18 at the preferred time of day. The rest of the participants did sport on average 86.7% of the preferred times. It can be expected that sport will mostly take place in the preferred period. For further studies, a clear time window should also be specified for the times of day (e.g. morning = 4 a.m. to 12 a.m., afternoon = 12 a.m. to 6 p.m., evening = 6 p.m. to 10 p.m. and night = 10 p.m. to 4 a.m.). Alternatively, specific time windows could be specified so that there is no overlap. Otherwise sport from 11 a.m. to 1 p.m. can be assigned to sport in the morning and sport at noon.
Furthermore, it was noticed that for people who do sport mostly in the evening, sport in the morning was always planned. On the other side, for the person doing sport mostly in the morning all her unplanned sport did happen in the evening. This provides an important clue when predicting the exact time of sports. For further studies, it would be advisable to additionally query the planned time when asked "whether you plan to do sport today".

4.9 Plan to Do Sport Today

Every morning the trial participants were asked whether they are going to do sport today. If the answer was *"yes"*, 75% of the time sport was performed on the same day. If the answer was *"maybe"* 20% of the time sport was performed on the same day. If the answer was *"no"* 7.6% of the time sport was performed on the same day, meaning that *"no"* applied in 92.4% of the cases. *"No"* therefore has the greatest significance for predicting sport followed by *"yes"*.

4.10 Unplanned Sport

It happened three times that a person spontaneously exercised. Furthermore, no differences between unplanned sport and planned sport were identified. On this basis, no statements can be made about sport prediction and this is probably not a useful indicator.

4.11 Sleep

Using the t-test, it was determined that the sleep quality entered in the morning of days with sport(A (average) = 3.51; SD (standard deviation) = 1.54) was better than on days without sport (A = 2.88; SD = 1.6). The difference was significant: t (129) = 2.3, p (alpha error) = 1.1%. Notably person 1, 2 and 4, which don't do sport in the morning, are standing out (see Figure 4.1). Thus, sleep in combination with the time of day is an important indicator for the prediction of sport. This is probably due to the fact of being tired in the evening when having bad sleep, which results in skipping the sport. In contrast, sleep quality seems to be only a small factor for people who do sport in the morning (person 3) and people who do sport at noon (person 5).

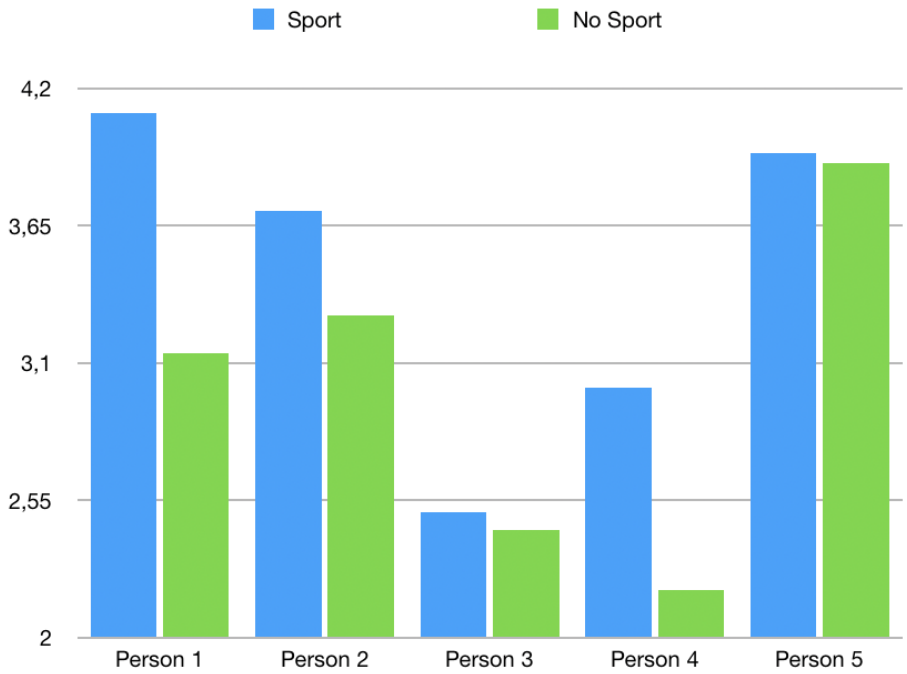


Figure 4.1: Average sleep rating per person on days with sport and without.

4.12 Weather

The study was conducted in January 2020. To evaluate this indicator, the average number of sport days was compared with the average number of sport days on days with rain and without rain. For this purpose, the individual sport times were considered and the weather checked at the respective times. The result is that on average 48.9% of the sport was done in the rain. In comparison, sport was generally performed on 47.7% of all days without considering the weather. This leads to the conclusion that there were no significant differences

in the rain. An exception to this, however, is person 2, who does four sports. When it was raining, person 2 exercised 38% of the time, but on days when it was not raining, person 2 exercised 64% of the time. Since person 2 is the only person who does outdoor sport (jogging), it makes sense that less sport takes place outdoor when it rains. The rest of the people only do sport at home or in the gym. According to the test persons, they drive to sport by car, whereby it is obvious that the weather plays a subordinate role here. Since the study suggests that weather only influences outdoor sports, a distinction should be made between outdoor and indoor sports. Accordingly, the question of how many outdoor sports are done could be added to the questionnaire. This should be related to the total number of sports and can be weighted accordingly. Since there was no differentiation between the sports, no probability can be assigned to this indicator. Person 1 did 26% less sport on rainy days, but only one sport out of four was an outdoor sport and it is not clear whether only the outdoor sport was affected by the weather. Accordingly, further research has to be done. Furthermore, the means of transport used could also be important for the influence of weather on sport and should be queried in future studies.

4.13 Satisfaction with Meals

Using the t-test it was determined that satisfaction with meals on days with sport (A = 3.87; SD = 1.85) were better than on days without sport (A = 3.37; SD = 0.87). The difference was significant: t (129) = 2.03, p = 2.2%.
Thus, satisfaction with meals is another indicator for predicting sport. In addition, it was checked whether after days when eating worse, the next day more likely sport is done. This was not the case.

4.13.1 Satisfaction with the First Meal

Using the t-test it was determined that the satisfaction with the first meal on days with sport (A = 3.82; SD = 1.02) were better than on days without sport (A = 3.21 ; SD = 1.14). The difference was very significant: t (76) = 3.21, p = 0.7%. The first meal is therefore another important indicator for predicting sport and is more significant than the remaining meals of the day.

4.14 Gender

A female subject reported that exercise was canceled due to the menstrual cycle. This means that a menstrual calendar could be created for women to increase the accuracy of the sport prediction for women. It should be noted that a single person has no significant informative value, but this is an indication that further studies could address. Depending on the target group under consideration, a pregnancy could also be taken into account, since this could also affect the sporting activity of women.

4.15 Further Possible Indicators

After questioning the subjects and evaluating the data, the following possible indicators for later studies emerged. These could be checked for their significance in further studies.

- **Disease:** A subject was unable to exercise for two weeks due to an illness. Injuries and surgery are also likely to result in less physical activity.
- **Holiday:** Holiday or vacation probably have an influence on the sporting activity.
- **Training Level:** Beginner athletes need a longer time to regenerate. They get sore muscles more likely for example. Accordingly, they probably do less sport.
- **Stress:** Psychological factors could play a further important role in relation to sport activity. For example, preparing for an exam or writing a thesis can change the sporting activity. On the one hand, sport could reduce stress, on the other hand, learning pressure could lead to a reduction in the sporting activity.
- **Diet:** Weight loss could increase sporting activity to lose more weight, whereas a calorie deficit could lead to energy loss, which could lead to less exercise.
- **Preferred Days:** Due to work or other reasons it can happen that a person does no sport on some days in the week. The other way round, it could be that due to working hours on Sunday there is an increased likelihood of sport.
- **Rehabilitation:** People with diseases such as rheumatism or pensioners in general offer their own target group, which should be considered separately, as sporting activity can be used for rehabilitation.

5 Prediction

The aim of the study was to find indicators with which sport can be predicted. In the following chapter, the results of the study are summarized in a app concept to predict sport. The aim of the concept is to be able to predict with a certain probability for the current day whether sport is done and in what time period this is likely to take place.
The study determined that the indicators

- **plan to do sport today**,
- **day of the week**,
- **number of sport units per week**,
- **sleep**,
- **satisfaction with meals, especially with the first meal** and
- **daytime**

affect whether a person does sport or not. When looking at these indicators, it is striking that they can be divided into two types of indicators. The indicators *satisfaction with meals*, the *plan to do sport*, *sleep* and the *number of sport units per week* provide information about **whether** a person is going to do sport. The *daytime* indicator can be used to tell **when** a person is exercising. The *day of the week* indicators provide both information about the "when" and "whether". For this reason, the prediction of sport is divided into **Whether-Prediction** and **When-Prediction**.

5.1 Whether-Prediction

In order to be able to predict sport, probabilities from 0% to 100% are assigned to the *Whether-Indicators*. The **average of the probabilities** is then formed. If the result is greater than 50%, it is assumed that sport is being done on the corresponding day. If it is less than 50%, it is assumed that no sport is being done. The values for the probabilities for the *Whether-Indicators* are chosen as follows.

5.1.1 Plan to Do Sport

For the entries *"yes"*, *"maybe"* and *"no"*, a list is created with zeros and ones. The ones stand for the fact that sport is done on that day. For each individual entry, the probability that

sport will take place on that day is calculated from the respective list based on the relative frequency of sport days. For example, if on two out of ten days with a *"maybe"* entry sport has been done, the probability that sport will take place on that day is 20%.

5.1.2 Day of the Week

For each day of the week a list is created with a sequence of zeros and ones. The ones stand for the fact that sport is done on that day. For each individual day of the week, the probability that sport will take place on that day is calculated from the respective list based on the relative frequency of sport days. For example, if on nine out of ten Mondays sport has been done, the probability that sport will take place on a Monday is 90%.

The following values of the probabilities are based on assumptions. The assumptions were made in order to be able to evaluate the basic idea of the concept.

5.1.3 Number of Sport Units Per Week

For this indicator, a list is saved for each week with the average number of sport units per week. From the current day, the number of days with sport is then compared with the average number of sport units per week. If in the past six days sport has been done below average, the probability is more than 50% that sport will be done today. If the current number is above average, the probability is below 50%. In both cases, the greater the deviation, the greater the difference is 50%. One way to assign a probability would be to change the probability by 20% for each full day of deviation (for half a day this means 10%). If sport is done two times in the past six days, but the average is three times a week the probability for this indicator is 80% that today sport will be done. The maximum value is 100% and the minimum value 0%.

5.1.4 Sleep

Since sleep only significantly influenced the sporting activity of the test subjects who exercised in the evening, this value is only considered for those who exercise in the evening. The sleep ratings are saved in a list. When entering the possible questions, the entered value is then compared with the average from the list. If the current day corresponds to the average, the probability for this indicator is 50%. For each deviation by one value, the probability is increased or decreased by 20%, depending on whether it is above or below the average value. The maximum value is 100% and the minimum value 0%.

5.1.5 Satisfaction with Meals

For the meal satisfaction the average meal satisfaction is stored in a list. For the current day, the average of the meals entered up to that point is compared. Thus, the first meal has the greatest influence on sport, which was determined in the study. If the current day corresponds to the average, the probability for this indicator is 50%. For each deviation by one value, the probability is increased or decreased by 10%, depending on whether it is above

or below the average value. The maximum value is 100% and the minimum value 0%. The increase or decrease rate was chosen by 10%, because the difference in average satisfaction with meals on sport days to non-sport days was less than in the sleep assessment.

5.2 When-Prediction

So far, the indicators have been used to determine whether sport is taking place on the current day. However, it is also interesting to know when this is likely to take place. In order to limit the time, the initial questionnaire asks at what time of the day sport is usually done.
With this information it can only be predicted that sport will take place in this period (morning, afternoon, evening or night).
To improve this statement, for each day of the week a list with the start and end times for each time of the day is saved (morning = 4 a.m. to 12 a.m., afternoon = 12 a.m. to 6 p.m., evening = 6 p.m. to 10 p.m. and night = 10 p.m. to 4 a.m.). The end time is saved to calculate the average sport length. The lists are then used to determine the average of the start times. The average is formed for each list of the times of the day. If there are several values for the average, a probability is assigned based on the relative frequency of values to which the respective average is assigned. For the end time, the average duration of a sport unit is calculated based on the start and end times and added up to the predicted start time.
Example 1: Sport has been done on three Mondays from 7:00 a.m. to 8:00 a.m., 6:00 a.m. to 6:30 a.m. and 8:00 a.m. to 10:00 a.m.
Every sport is done in the morning.
The average start time is 7:00 a.m.
The average length of a sport session is 70 minutes.
It is therefore assumed that if sport takes place on a Monday, it will take place from 7:00 p.m. to 8:10 p.m. in the morning.
Example 2: Sport has been done on four Saturdays from 10 a.m. to 11 a.m., 8 p.m. to 9 p.m., 9:10 p.m. to 10 p.m. and 3 p.m. to 4 p.m.
One sport is done in the morning, one in the noon and two in the evening.
The average start times are 10:00 a.m. (morning), 8:35 p.m. (evening) and 3:00 p.m.(afternoon)
The average length of a sport session is 57.5 minutes.
It is therefore assumed that if sport takes place on a Saturday, the probability that sport will take place in the morning from 10:00 a.m. to 10:57 a.m. is 25%, the probability that sport will take place in the afternoon from 3:00 p.m. to 3:57 p.m. is 25% and the probability that sport will take place in the evening from 8:35 p.m. to 9:32 p.m. is 50%.

5.3 Evaluation of the Prediction Concept

In order to evaluate the accuracy of the prediction concept, this concept was applied based on the data from the last week of the study. It was checked on how many days sport would be correctly predicted. As a result, the concept correctly predicted in 84.4% of the days whether sport was taking place or not. In addition, the start time differs by an average of 1 hour and 19

minutes from the predicted time. Since only data for a period of four weeks are available for testing the concept, it can be expected that the concept can predict sport even more precisely over a longer period. Furthermore, as a result of further research, more indicators can be added into the concept or indicators already used can be weighted differently in order to make the prediction even more precise.
The weaknesses of the concept is in predicting sport for people who rarely do sport, always do sport on other days and times and often wrongly claim that they are doing sport today. This concept does not cover several sport units in one day. Even illnesses or the start of a new sport or in general the immediate change in sporting activity can be difficult to predict with this concept.

6 Outlook

This study tried to answer the question of the extent to which sport can be predicted. For this purpose, a four-week study was carried out using an app developed for this purpose.
The result of the study shows that the **plan to do sport today**, the **day of the week**, the **number of sport units per week**, the **sleep**, the **satisfaction with meals, especially with the first meal** and the **daytime** are indicators to predict sports. These indicators were then summarized in a concept for a sport prediction app. An evaluation of the concept showed that in the last week of the study it would be predicted correctly with an accuracy of 84.4% whether sport would have been done or not.
Due to the small number of participants (n = 5), no significant statements can be made about the general public, but interesting observations could be made, but mechanisms can be found. These mechanisms can then be checked in general validity in further studies.
Several aspects should be checked in these follow-up studies. At the beginning it should be checked whether the prediction indicators are also valid for other sports. Then the indicators from Chapter 4.15 (Further Possible Indicators) should also be checked. In the study of this bachelor thesis, the focus was on strength training. Furthermore, a long-term study with a larger number of test participants should be considered to make statements about the general public.
For further research, an app has to be implemented that contains the functions of the concept. The already implemented app can be expanded for this. Since the app was only implemented for Android, a new app for iOS could be implemented. A platform-independent implementation can also be considered. If a native implementation is chosen, it makes sense to integrate the use of smartwatches.
Finally, this bachelor thesis should provide a basis for the research of the sports and health scientists at the Technical University of Munich. The results should help to research whether sporting activity has an influence on the choice of your own meals and whether you eat more appropriately if you commit to your diet before exercising.

Bibliography

[1] B. für Wirtschaft und Energie (BMWi). *Sportwirtschaft – Fakten & Zahlen*. Online available on: https://www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/sportwirtschaft-fakten-und-zahlen.pdf%3F__blob%3DpublicationFile%26v%3D12; accessed on 30.1.2020. April 2018.

[2] K. Koehler, J. B. Boron, T. M. Garvin, M. R. Bice, and J. R. Stevens. "Differential relationship between physical activity and intake of added sugar and nutrient-dense foods: A cross-sectional analysis". In: *Appetite* 140 (2019), pp. 91–97.

[3] K. Koehler, M. J. De Souza, and N. Williams. "Less-than-expected weight loss in normal-weight women undergoing caloric restriction and exercise is accompanied by preservation of fat-free mass and metabolic adaptations". In: *European journal of clinical nutrition* 71.3 (2017), pp. 365–371.

[4] *Motion sensors | Android Developers*. Online available on: https://developer.android.com/guide/topics/sensors/sensors_motion; accessed on 30.1.2020.