Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

UDC 50.43 

Azamat Serek

# Analysis of data to improve system of an educational organization

THESIS

Presented in Partial Fulfillment for the
Degree of Master of Science in Computing Systems and Software
(degree code: 6M070400)
Department of Computer Sciences
Faculty of Engineering and Natural Sciences

Supervisor: **Meirambek Zhaparov**

Kaskelen, 2020

# Abstract

In this thesis, there was done a set of computational experiments on the datasets of educational organizations. In the first part of the work, there were executed lots of experiments with decision tree ID3 algorithm on the educational camp dataset ("Educon") that automatically predicts a participant's feedback using Scikit-learn library and extrapolatory data analysis of that was done using Pandas library and Python programming language. The experimental results showed that the most optimal maximum depth (which is the number of edges starting from the root till the leaf) for the decision tree is 3 and the most optimal minimum number of splits (which is the minimum amount of samples of the dataset that are required to split an internal node of the decision tree) is 192. Based on that, there was achieved optimum results of precision, recall, and f1 score machine learning metrics that vary between 75 to 98 depending on the change of tunable variables of the ID3 algorithm. In the subsequent parts of the thesis, the information extraction system was built based on an educational camp dataset and recommendations for hackathon improvement were derived. The datasets are not open-source and were collected manually through the use of surveys.

# Аңдатпа

Диссертациялық жұмыста білім беру ұйымдарының мәліметтер базасында бірқатар тәжірибелік эксперименттер жүргізілді. Жұмыстың бірінші бөлігінде білім беру лагерінің мәліметтер базасында (" Educon ") ID3 шешім ағашының алгоритмімен көптеген эксперименттер орындалды, бұл қатысушының пікірін Scikit-learn кітапханасы мен экстраполяторлық деректерді талдау көмегімен автоматты түрде болжайды. Pandas кітапханасы мен Python бағдарламалау тілін қолдану. Эксперименталды нәтижелер көрсеткендей, шешім ағашына арналған ең оңтайлы максималды тереңдік (бұл тамырдан бастап жапыраққа дейінгі жиектер саны) 3 және бөлудің ең оңтайлы минималды саны (бұл мәліметтер жиынтығы үлгілерінің ең аз мөлшері) шешім ағашының ішкі түйінін бөлу үшін қажет 192). Осыған сүйене отырып, дәлдікті, еске түсіруді және f1 баллдық машинаны оқудың өлшемдерінің оңтайлы нәтижелеріне қол жеткізілді, олар өзгермелі айнымалы шамалардың өзгеруіне байланысты 75-тен 98-ге дейін өзгереді. ID3 алгоритмі. Диссертациялық жұмыстың келесі бөлімдерінде ақпараттық лагерьдің базасында ақпараттарды жинау жүйесі құрылды және хакхатонды жақсарту бойынша ұсыныстар алынды. Мәліметтер жиынтығы ашық емес және сауалнаманы қолдану арқылы қолмен жиналған.лжау.

# Аннотация

В этом тезисе был проведен ряд вычислительных экспериментов на наборах данных образовательных организаций. В первой части работы было выполнено множество экспериментов с алгоритмом дерева решений ID3 в наборе данных образовательного лагеря (" Educon "), который автоматически предсказывает обратную связь участника с помощью библиотеки Scikit-learn и был проведен экстраполяционный анализ данных. используя библиотеку Pandas и язык программирования Python. Результаты эксперимента показали, что наиболее оптимальная максимальная глубина (то есть число ребер, начиная с корня до листа) для дерева решений составляет 3, а наиболее оптимальное минимальное количество расщеплений (которое является минимальным количеством выборок набора данны,). которые необходимы для разделения внутреннего узла дерева решений) - 192. Исходя из этого, были достигнуты оптимальные результаты метрик машинного обучения точности, отзыва и оценки f1, которые варьируются от 75 до 98 в зависимости от изменения настраиваемых переменных алгоритм ID3. В последующих частях диссертации система извлечения информации была построена на основе набора данных образовательного лагеря и были получены рекомендации по улучшению хакатона. Наборы данных не являются открытым исходным кодом и были собраны вручную с помощью опросов.

# Acknowledgements

Thanks to my thesis supervisor for constant support and useful discussion. Thanks to the external reviewer for very useful feedback that helped to significantly improve the current work. Thanks to my parents and brothers for all kinds support.

# Contents


**1 Introduction** **7**

1.1 Motivation . . . . . 7

1.2 Aims and Objectives . . . . . 9

1.3 Thesis Outline . . . . . 9

**2 Preliminaries** **10**

2.1 Background . . . . . 10

2.1.1 Educational association overview . . . . . 10

2.1.2 Overview of the particular instructive associations . . . 10

2.2 Preliminaries for the third study . . . . . 11

**3 Related work** **15**

3.1 The first and the second study . . . . . 15

3.1.1 Overview of explicit instructive camps . . . . . 15

3.2 The third study . . . . . 16

3.3 The fourth study . . . . . 18

**4 The first study** **19**

4.1 Objectives . . . . . 19

4.2 Limitation . . . . . 19

4.3 Justification of the picked methodology . . . . . 19

4.4 Used technologies . . . . . 20

4.5 Analysis of "EduCon" spring camp . . . . . 20

4.6 Recommendations . . . . . 30

**5 The second study** **31**

5.1 Objectives . . . . . 31

5.2 Limitations . . . . . 31
5.3 Justification of the picked methodology . . . . . 31
5.4 Used technologies . . . . . 32
5.5 Results . . . . . 37

**6 The third study** **40**
6.1 Methods and materials . . . . . 40
6.1.1 Dataset description . . . . . 40
6.1.2 Used technologies . . . . . 40
6.1.3 Inner mechanics of the system . . . . . 40
6.1.4 Justification of the picked methodology . . . . . 41
6.1.5 Limitations . . . . . 42
6.2 Results and discussion . . . . . 43

**7 The fourth study** **45**
7.1 Overview . . . . . 45
7.2 Methods and materials . . . . . 47
7.2.1 Dataset description . . . . . 47
7.2.2 Used technologies . . . . . 47
7.2.3 Justification of the chosen methodology . . . . . 47
7.3 Results and discussion . . . . . 49
7.3.1 Exploratory data analysis . . . . . 49

**8 Conclusion** **58**

**References** **59**

# 1. Introduction

## 1.1 Motivation

Educational camps that take place in summer become observable wherever all through the world [1]–[3]. In order to improve the potential results of the planning camp affiliation, we study the data of the spring camp in Kazakhstan. We are endeavoring to react to the subject of how the readiness framework can be improved the reason of a formally feasible research structure - EduCon. The investigation that empowered us was a way to deal with improve it. Concerning general hugeness of the topic, the issue of improving preparing is as yet prominent in the audit requests of the latest five years, as showed up by the data on Google Trends [4].

Data collection systems are used to dispose of significant information or to structure data that is semi-key in nature or non-essential in nature. Starting at now, data variety procedures have various businesses. [5]–[9] Data cancellation systems can benefit relationship by orchestrating data in a composed, easy to-use course and to get data on rough data.

In any case, getting ready camps need data variety structures. As such, they lose the recently referenced points of interest of asserting data variety structures.

Getting ready camps are persistent informative and unwinding practices for youths that commonly last about a month. They are notable around the globe. [10]–[14] Learning camps help individuals with getting data, cooperate with people, and give fragile capacities.

In this work, a data variety structure depended on a starting at now adequately settled getting ready camp - the readiness guiding association 'EduCon' '.

'EduCon' is an availability and coordinating association with a couple of broad stretches of time connected with making workplaces as a test association for stu-

dent general preparing sub-inspect. The association has been worked for over 20 years by a gathering of qualified authorities invest critical energy in planning. The alliance sets up pilot associations for informational assignments and continually improves the possibility of the things publicized. In test associations, drawings and diagrams are used in student studies to fathom and drive the endeavor. Along these lines, " EduCon " will work personally with Dostyk to focus on the explanation of the comprehension and give its association to optional schools and instructing needs. [15]

The inalienable system licenses you to intensely watch the individual information of the individuals in each camp with the looking at test results and access the ordinary test results each month for each city while imagining in a straightforward manner. The work can help planning with outside pioneers improve camp organization and make the basic changes subject to the results got.

Hackathons are standard all around the globe [16]–[18]. They are express activities that normally last around 24 hours and in which gatherings are molded to present their innovative responses for given IT issues. This article examined the outcomes of the Paragon Hackathon review, which was guided from 10 to 11. January 2020 at Paragon International University.

The ParagonIU Hackathon was a 2-day coding design that incited individuals to develop a web application using the Facebook API.

The work asks what legitimate changes ought to happen to also develop the hackathon. The results gained can possibly be extrapolated to various hackons, especially in Asia, due to the higher probability of closeness of individuals.

Paragon International University (some time back known as Zaman University) was worked in 2010 as the second unit of Zaman's indicating affiliations (by and by the Paragon Education Company). Zaman University was remade and was formally seen as the International University of Paragon in a sub-introduction gave by the Royal Government of Cambodia on 28 January 2019.

By its eminence, Paragon International University intends to be the most noteworthy school in Cambodia, showing enormity in research and web the administrators, giving schools exhibited endeavors and making a marvelous HR bunch that will extend the general progression and improvement of Cambodia and its kinfolk. In understanding the overall world and the enthusiasm for overall correspondence, the language of direction at Paragon International University has

been English from the soonest beginning stage.

## 1.2 Aims and Objectives

The purpose of the work is to improve informational affiliations.

There are the going with objectives:

- Collect data from "EduCon" about Spring Camp and Winter Camp
- Analyze the recently referenced datasets using Python, Pandas
- Write recommendation motivations behind advancement reliant on the recently referenced extrapolatory data assessment
- Apply id3 figurings to foresee contribution of a part
- Analyze decision trees figurings to find the most suitable one for the task
- Create an information extraction system on google sheets subject to EduCon data
- Build customized data portrayal for the recently referenced instrument
- Analyze Paragon Hackathon and make proposals motivations behind upgrades subject to that
- Define the effect of sex in the audit of Paragon IT Hackathon

## 1.3 Thesis Outline

The essential area is Introduction part. It is this one that you are starting at now examining. It gives understanding into the work done. In Chapter 3 we review related work and plan the issue to clarify. Area 4 is depicting the response for the issue of progress of EduCon camp reliant on their dataset. In The second study segment there delineated an execution of customized figure of information assessment of a part for EduCon camp subject to ID3 decision trees estimation. In The fourth study information extraction structure for EduCon camp is explained, and in Chapter 4 there portrayed a data assessment of the Paragon IT hackathon to find the techniques for its improvement.

# 2. Preliminaries

## 2.1 Background

In this area, a few foundations to the peruser will be introduced to place him in the correct setting. In the first place, we give a thought regarding instructive associations, at that point we present some particular camps which will be worried by our methodology.

Figure 2.1 shows the techniques and materials of the proposal.

### 2.1.1 Educational association overview

The instructive association is an association contained inside the extent of training, which can be considered as an organization offering instructive types of assistance.

Favorable circumstances of instructive associations

- Educating people group
- Providing administrations for learning
- Contributing to the advancement of the human psyche

### 2.1.2 Overview of the particular instructive associations

- Educational camps
  - Their motivation is to provide for the members of the camp instructive and social improvement in a supporting and thriving condition.
- Universities

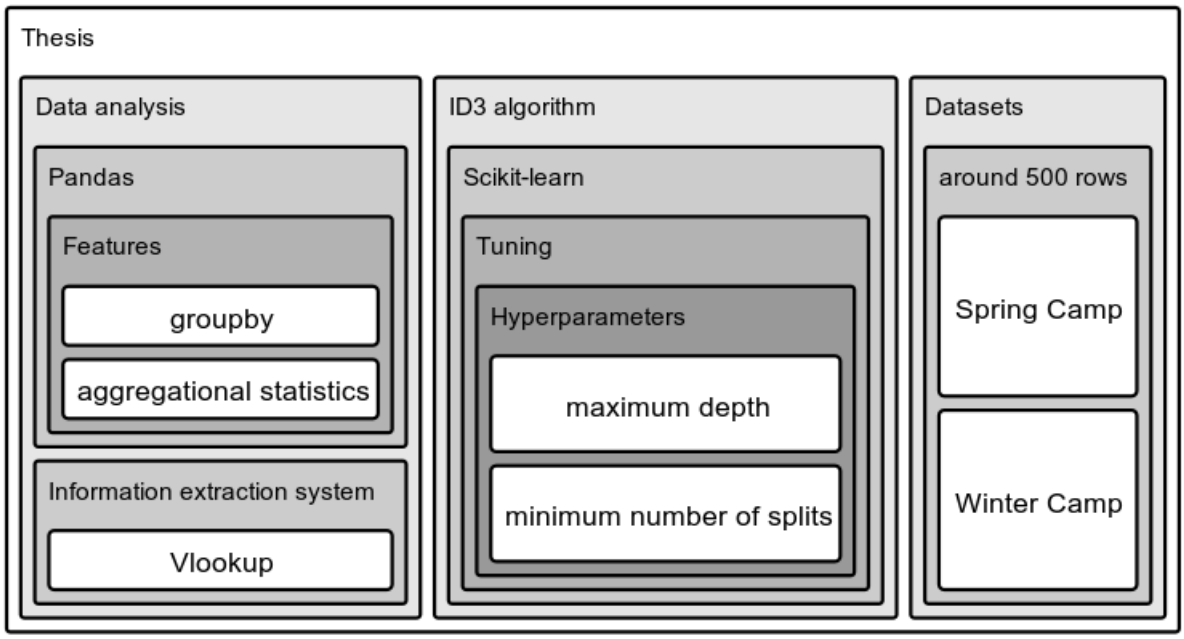


Figure 2.1: Method

– Their motivation is to give significant level training and degrees to understudies, and where scholastic research is directed.

## 2.2 Preliminaries for the third study

Applying information extraction to content is identified with the issue of improving substance to make sense of everything point of view on the information in free substance. The general objective is to make the more amazing substance of the machines powerfully useful in sentence managing. The creation line's IE affiliations and subtasks are:

- Setup Fill: Unpacking a fixed field plan, for example, vindicated liable gatherings, lamentable mishaps, time, etc in a Forth paper article about the dread challenger Assault.

- Event Extraction: When there is a record, passes on at any rate zero particular ways Layouts. For instance, a paper article can portray different assaults by dread challengers.

- Database individuals: Fill in a database of fortresses that have been given a great deal of records. Continually, the database is triple (substance 1, connection, section 2), for instance (Barack Obama, amigo, Michelle Obama)

- Named Acknowledgment: Acknowledgment of perceived part names (for people and affiliations), the titles of places, transient explanations, and wonderful numerical verbalizations utilizing existing information in the territory or information exonerated from different sentences. For the most part, the statement task fuses doling out one sort of identifier to the confined portion. An even more away from is the obvious confirmation of substances that consideration on the specific fragments without having noteworthy data on central conditions. For instance, in setting up the explanation "M. Smith likes from a point," the district of a fragment called would reason confirmation that the enunciation "M. Smith" recommends an individual, at any rate without having (or using) any information on a specific M. Smith who is (or " perhaps ") the individual to whom the sentence intimates.
- Coreference objectives: the locale of center focuses and anaphorical connection between content sections. In IE affiliations, this is ordinarily obliged to discovering connection between heretofore existing named portions. For instance, "For the most part Business Machines" and "IBM" recommend a comparable genuine substance. In the event that we take the two sentences "Yerbol likes cycling. Anyway, he couldn't ponder cycling," it legitimizes seeing that "he" infers the starting late perceived unit "Yerbol."
- Relationship picking: indisputable check of, for instance, relationship between substances,
  - Individual work for an alliance (kept from " Yerbol works for some company ")
  - Person in LOCATION (expelled from " Yerbol is in Italy ")

- Semi-shaped information extraction which may prescribe any IE that attempts to restore a sort of information structure that has been lost through scattering.
  - Table extraction: finding and ousting tables from various reports.
  - Extraction of information from tabular data: removing information in a filtered through route from the tables. This is a more dazzling

assignment than table extraction, as table extraction is only the key turn of events, while understanding the organizations of the cells,

  - Remarks extraction: separating remarks from the confirmed substance of article to reestablish the relationship between a writer of each sentence

- Language and language assessment

  - Phrasing extraction: finding the important terms for a given corpus that can vary in size

- Extraction of audio materials

  - music extraction: finding important trademark in a sound sign taken from a given combination; for instance time game plans of events of percussive sounds can be ousted to address the essential cadenced bit of a music piece.

Note that this rundown isn't thorough and that the unmistakable significance of IE rehearses isn't routinely perceived and that different strategies join diverse sub-assignments of IE to accomplish a continuously wide objective. Man-made reasoning, quantifiable appraisal or potentially trademark language managing are reliably utilized in IE.

IE on non-content reports is changing into an obviously captivating topic[when?] in research, and data hidden and sound records can now[when?] be passed on in a basic level structure as it is done on content. This commonly prompts the mix of expelled data from various sorts of records and sources.

There are numerous nuances and complex methods in the information assortment process, however a decent beginning for a learner is to recollect:

Content in and information out

To refine this moderate method of portraying information assortment, the procedure includes changing unstructured content or an assortment of writings into realities (i.e., formal, machine-discernible explanations of the sort "Yerbol is the writer of this material") that are as yet occupied (filled) in a database, (for instance, the some database of literature). ).

Regularly, unstructured writings give organized data on the accompanying fundamental undertakings:

Content Preprocessing - This content sets you up for handling utilizing computational phonetic apparatuses, for example, tokenization, sentence division, morphological examination.

Finding and arranging ideas - here the kinds of individuals, things, spots, occasions, and other predefined ideas are referenced and grouped.

Consolidating ideas - this is utilized to distinguish the connections between the ideas picked.

Total - this sub-task concerns the introduction of the removed information in a standard configuration.

Dispose of commotion - this subtask requires the expulsion of copy information.

Database enhancement - now, the extricated data is gone into the database for additional utilization.

Information assortment can be completely robotized or performed with human help.

Regularly, the best information deletion arrangements are a mix of computerized strategies and human preparing.

# 3. Related work

## 3.1 The first and the second study

Summer preparing camps become unmistakable all over the world [1]–[3]. So as to improve the conceivable outcomes of the preparation camp association, we study the information of the spring camp in Kazakhstan. We are attempting to respond to the topic of how the preparation system can be enhanced the premise of an officially successful research structure - EduCon. The exploration that empowered us was an approach to improve it. With respect to general significance of the point, the issue of improving instruction is still notable in the overview inquiries of the most recent five years, as appeared by the information on Google Trends [4].

### 3.1.1 Overview of explicit instructive camps

In [19] proposed teaming up with outer associations to give the mid year

innovation camps for secondary school understudies, for example, science historical centers, mentoring of philanthropies, and so forth. The educational plan is worked in a joint effort with educators with associations. After the camp finished, the association got criticism from understudies, their folks, and teachers. The criticism was certain, which prompted an expansion in understudy intrigue and guide impetuses.

In [20] an issue was found to bring down enthusiasm for figuring programs that could devastatingly affect American security. The creators proposed seven days in length day camp program for secondary school understudies that gave them openings and openings in the field of data innovation. As indicated by the input assembled from the members, the camp expanded their enthusiasm for data innovation and improved their inspiration to take part more in data innovation/innovation courses.

In [21] the motivation behind why the quantity of ladies going to "coding fire up camps" is generally higher contrasted with ladies taking part in the college program in data innovation (35–14%) by gaining from foundation 9 ladies who partook in two diverse camp conditions was analyzed. The creators 'tests affirmed that a potential reason for the above issue is the impact of a male-commanded culture.

## 3.2 The third study

Information assortment frameworks are utilized to eliminate significant data or to structure information that is semi-fundamental in nature or non-basic in nature. Right now, information assortment strategies have numerous employments. [5]–[9] Information extraction frameworks can profit associations in sorting out data in an organized simple to-utilize path and for information extraction from crude information.

Be that as it may, preparing camps need information assortment frameworks. In this way, they lose the above advantages by having information deletion frameworks.

Preparing camps are continuous instructive and relaxation exercises for youngsters that generally last about a month. They are exceptionally well known around the world. [10]–[14] Learning camps assist members with picking up information, team up with individuals, and give delicate abilities.

In this work, an information assortment framework was based on an as of now effectively settled preparing camp - the preparation counseling organization 'EduCon' '.

'EduCon' is an arrangement and guiding organization with a few significant stretches of time associated with making offices as a test organization for undergrad general training sub-investigate. The organization has been worked for more than 20 years by a group of qualified specialists represent considerable authority in preparing. The affiliation sets up pilot organizations for instructive ventures and constantly improves the idea of the items advertised. In test organizations, drawings and graphs are utilized in undergrad studies to comprehend and drive the errand. So also, " EduCon " will work intimately with Dostyk to concentrate on the reason of the understanding and give its organization to discretionary schools

and instructing needs. [15]

The manufactured framework permits powerfully observe individual data of each camp member with comparing test results and gain admittance to average test results every month for every city envisioning in an easy to use way. The work can help directors in instructive camps to improve control of a camp and roll out vital improvements dependent on acquired outcomes.

## 3.3 The fourth study

In [22] there was sorted out Think Global Hack Local (TGHL) hackathon, that associated non-benefit associations with understudy designers to address the issue of females not partaking hackathons with a similar recurrence as guys by making the hackathon non-serious and network based domain.

In [23] there was led a college hackathon to battle intestinal sickness and zika with rambles. It was a 1-day occasion at California State University, where a great deal of members dealt with the tasks to battle mosquito-borne sickness with land-based and flying automatons.

In [24] there was sorted out Oak Ridge Leadership Compute Facilities GPU Hackathon. It was an inventive and creative system to cover GPU programming in the domain of hackathons.

In [25], there was indicated that a hackathon is a promising methodology in programming building because of an expansion in the pertinence of speed in programming designing.

In [26], a hackathon procedure was applied for college class ventures. Studies demonstrated that learning is one of the principle sparks of hackathon members, smoothing out the ideation procedure with customary expectations in brief time-frame outlines prompts target conversation and snappy dynamic.

In all the previously mentioned hackathons, the reactions of the degree to what the members enjoyed hackathons and felt the expansion of information was for the most part positive. It likewise shows that the hackathon has its utilization in such a large number of areas of life.

# 4. The first study

## 4.1 Objectives

Explore the nature of instruction and social fulfillment at the "EduCon" camp and get proposal focuses for development.

## 4.2 Limitation

The objective gathering was Kazakh understudies, age bunches 10-18 years of age, who took an interest in the overnight camp 'EduCon' 'in the spring of 2018. The examination materials comprised of an inventive outdoors program and a paper-based survey group. The review comprised of inquiries on social and learning fulfillment with their assessment of the quarters, most loved/disliked subject, nourishment, exercises, occasion, and other related points. In the wake of gathering the information, it was physically moved to Google Sheet. Information were investigated by content examination strategy, means, and mode esteems.

## 4.3 Justification of the picked methodology

The survey was in paper structure in view of the chance of not realizing how to round out an online poll for all kids or not having the important gear. The issue was worried about the particulars of the camp as it assists with pulling together on solid regions of the camp.

Table 4.1: Quantity of participants of the camp

| City | Number of students |
|---|---|
| Aktau city | 13 |
| Aktobe city | 22 |
| Almaty city | 53 |
| Astana city | 61 |
| Atyrau city | 32 |
| Qaragandy city | 19 |
| Qokshetau city | 12 |
| Qostanai city | 6 |
| Qyzylorda city | 20 |
| No Name | 4 |
| Oral city | 13 |
| Osqemen | 14 |
| Pavlodar | 13 |
| Shymqent | 23 |
| Taraz | 20 |
| total | 325 |

## 4.4 Used technologies

To achieve the previously mentioned reason Python, and Advanced Google Sheets have been used.

## 4.5 Analysis of "EduCon" spring camp

In this section, we will show an undeniably clear depiction of the examination that has been done on the dataset.

Table 4.1 records the city names and the amounts of understudies that went from the urban territories. Kazakhstan is the greatest country in Central Asia. The understudies went to this camp from 14 urban networks, all around the country.

Table 4.2 records the decision of the sub-study profile items. Most of the topics collected were materials sciences and arithmetic, with 150 sub-studies out of 325 selected topics. This is 46 percent of the undergraduate studies. Different

Table 4.2: Type of taught subject

| Subject name | Number of students |
|---|---|
| Biology and Chemistry | 63 |
| Biology and Geography | 1 |
| English and Geography subject | 8 |
| English and History | 58 |
| History and Human. Society. Law. | 11 |
| Kazakh language subject and History | 8 |
| Mathematics and Geography | 24 |
| Physics and Mathematics | 150 |
| No response | 2 |
| total | 325 |

subjects were picked with a much more modest number of undergraduate studies: science and science 63 undergraduate studies, English and history 58 undergraduate studies, geology and arithmetic 24 undergraduate studies, the remaining topics were selected with less than 10 subgraduate studies. Based on this knowledge, it is possible to understand that science and materials science are the most in demand.

Table 4.3 record the propensity to maintain inadequate camp studies. Lower studies made 119 decisions in favor of keeping the whole food. 50 sub-researchers prefer food prepared for lunch, for example pilaf (27), chicken (17) and soup (16). 22 and 21 votes chose the dish prepared for dinner and breakfast separately. Use this table to understand what types of foods are consumed the most.

Table 4.5 records the tendencies of underestimations in mass stimulation. It can be seen from the table that the best events were the Bye-Bye celebrations, which were chosen by 99 undergraduates. In addition, 53 undergraduates preferred football, with slightly fewer undergraduates (43) choosing different courses. The most amazing result of all the spring camp situations shows the abnormal state of readiness of the coordinators.

Table 4.4 records ominous exercises. 10 sub-studies gave a bad assessment of the situation "What? At what? At what stage?" And likewise, 7 students chose Nauryz. These insights help coordinators filter out mass briefings. By taking advantage of it, you can, without a large part, find the tendencies of sub-studies. Coordinators can banish situations that have not been further explored

Table 4.3: The meal which they liked the most

| Food name | Number of students |
|---|---|
| All | 119 |
| bayursak meal | 2 |
| meal for the breakfast | 21 |
| buckwheat meal | 3 |
| chicken burger | 17 |
| delicious cofte | 8 |
| dinner meal | 22 |
| fishery food | 3 |
| doner | 10 |
| lunch meal | 50 |
| meat related food | 5 |
| delicious pilaf | 27 |
| porridge | 10 |
| puree with cutlets | 12 |
| soup with fish | 16 |
| total | 325 |

and add exercises that are least focused or not placed together in any part of the imagination. Along these lines, they construct the nature of events.

In Table 4.6, it shows the tendencies in the camp's undergraduate subjects. From the votes collected, you can see that the most popular topic was history, which received 63 ballots. In addition, 27 votes were cast for the science topic and 23 votes for the science topic. This table shows that the tendencies of elementary students are exceptional. inclinations are extraordinary. Table 4.7 ecords the conclusions of the research results on the issues discussed in the camp. Lower researchers chose the topics they liked the least. Above all, votes were cast on the topic of numerical education (26). In addition, the subject of materials science was seriously evaluated by 17 sub-studies. According to 61 sub-surveys, they did not have such a dissatisfied topic. This table helps to improve the nature of the guided topics and to create the conditions for identifying the benefits of undergraduate studies.

Table 4.8 records the students' assessment of the apartment or rather its most enjoyable meetings. We can see that more votes were cast (72) along the way, that all conditions are good. The cleanliness of the room is one of the great

Table 4.4: The meal which they did not like the most

| Title | Quantity of people |
|---|---|
| Cinema related entertainment | 1 |
| Dancing | 7 |
| They were not participants | 2 |
| Do tasks related to the studies | 4 |
| Soccer | 3 |
| IQ games | 5 |
| Nauryz entertainment | 7 |
| Phototaking activities | 6 |
| Playing video games | 4 |
| Participating on the seminars | 9 |
| Party devoted to freshment | 3 |
| Intellectual game | 10 |
| Everything is good game | 60 |
| No response | 204 |
| total | 325 |

Table 4.5: The loathed activity

| Activity | Number of students |
|---|---|
| «Photoshooting» | 24 |
| All | 22 |
| Anti-kafe | 6 |
| Bye Bye Party | 99 |
| Movie entertainments | 6 |
| Was not participant | 3 |
| Dancing | 2 |
| Soccer | 53 |
| Nauryz entertainment | 4 |
| PlayStation4 | 8 |
| Seminar | 46 |
| Sport's event | 5 |
| Tennis related games | 7 |
| Party for freshman | 10 |
| IQ game | 8 |
| No response | 22 |
| total | 325 |

Table 4.6: Favorable subject

| Taught subject | Quantity of people |
|---|---|
| All | 47 |
| Biology | 9 |
| Chemistry subject | 8 |
| Chemistry and Biology | 23 |
| English | 12 |
| English, History | 12 |
| Geography | 6 |
| Geography and Mathematics | 3 |
| Geography, History | 5 |
| History | 63 |
| History, Biology | 6 |
| History, Mathematics | 16 |
| History,Physics | 10 |
| Kazakh language, Literature | 7 |
| Mathematics subject | 27 |
| Mathematical literacy | 20 |
| Physics | 24 |
| Physics, Mathematics | 17 |
| Reading literacy | 4 |
| No response | 6 |
| total | 325 |

Table 4.7: Loathed taught subject

| Subject | Number of students |
|---|---|
| Chemistry | 3 |
| English | 2 |
| Geography | 1 |
| History | 11 |
| Law | 1 |
| Mathematical literacy | 26 |
| Mathematics | 7 |
| NO such | 61 |
| Physics | 17 |
| Reading literacy | 12 |
| No response | 184 |
| total | 325 |

Table 4.8: What they liked in the dormitory

| The good side of dormitory | Number of students |
|---|---|
| Different Activities | 4 |
| All conditions | 71 |
| As home | 8 |
| Beautiful | 4 |
| Big, comfortable | 10 |
| Cleanery | 24 |
| Convenient | 41 |
| Interior | 9 |
| Friendships | 4 |
| I wasnt a resident | 9 |
| room for cleaning | 8 |
| good | 6 |
| Tennis related games | 9 |
| WiFi | 2 |
| Clean rooms | 58 |
| No response | 62 |
| total | 325 |

Table 4.9: What they did not like in the dormitory

| The bad side of dormitory | Number of students |
|---|---|
| Bad WI-FI | 11 |
| Can't cook. | 3 |
| Coldly | 7 |
| Dirty | 3 |
| I didn't live | 3 |
| looks like a labyrinth | 2 |
| no minuses | 52 |
| no mirrors | 11 |
| Rooms don't close | 8 |
| Small shower and little cabinets | 17 |
| Time limit 06:00 to 23:00 | 4 |
| very noisy | 7 |
| no response | 207 |
| total | 325 |

Table 4.10: ”Spring camp” satisfied you?

| Answer | Number of students |
|---|---|
| 50/50 | 5 |
| No | 8 |
| Yes ! Everything was perfect | 304 |
| no response | 8 |
| total | 325 |

Table 4.11: Food rate

| Food rate | Number of students |
|---|---|
| 4 | 2 |
| 5 | 1 |
| 5 | 10 |
| 6 | 16 |
| 7 | 44 |
| 8 | 66 |
| 9 | 47 |
| 10 | 143 |
| no response | 1 |
| total | 325 |

Table 4.12: Dormitory rate

| Dormitory rate | Number of students |
|---|---|
| 0 | 1 |
| 1 | 1 |
| 3 | 1 |
| 5 | 5 |
| 6 | 5 |
| 7 | 16 |
| 8 | 33 |
| 9 | 60 |
| 10 | 179 |
| no response | 33 |
| total | 325 |

Table 4.13: Subject rate

| Subject rate | Number of students |
|---|---|
| 3 | 1 |
| 4 | 1 |
| 5 | 3 |
| 6 | 4 |
| 19 | 18 |
| 8 | 46 |
| 9 | 70 |
| 10 | 184 |
| no response | 1 |
| total | 325 |

Table 4.14: Event rate

| Event rate | Number of students |
|---|---|
| 1 | 1 |
| 2 | 3 |
| 3 | 2 |
| 5 | 10 |
| 8 | 10 |
| 7 | 22 |
| 8 | 39 |
| 9 | 49 |
| 10 | 183 |
| no response | 12 |
| total | 325 |

aspects of the quarters, which received 58 ballots. 42 undergraduates found the accommodation to be a decent half of the quarter. These insights will help you feel the quality of the quarters and the comforts, their upper and lower parts, through the eyes of the undergraduate.

Table 4.9 records the students' assessment of living, or rather its most desirable meetings. We can see that 53 basic research voiced that there is no terrible side to living. 10 area research did not care that there are no mirrors in the room; a similar number of sub-studies did not care about the small shower, and as evidenced by the 11 sub-studies of the quarters, terrible Wi-Fi has been seriously obtained. With this measurement, you can learn about the disadvantages of residence and further dispose of them for the placement of undergraduate studies.

Table 4.10 stores the answers to the query "Did you fill the spring camp?" According to 304 sub-surveys, all was well. This shows that everything is more than well composed and that it meets the aspirations of the undergraduate studies.

Table 4.11 stores an assessment of the maintenance of sub-studies in the camp. Similarly, we can see that 143 sub-studies out of 325 evaluated nutrition 10. 44 sub-studies were evaluated with 9, 66 sub-studies with 8 points, 42 sub-studies with 7 points. Children under 6 years of age were evaluated in 29 substudies. As this data shows, we can understand that the camp was excellent and many liked it.

Table 4.12 stores an estimate of camp sub-study housing. The aftermath of the review shows that of the 174 sub-studies, 325 rated Lodging 10 points as the most surprising and complete. 60 sub-studies were evaluated with nine, 31 field studies with 8 points, 27 field studies with 7 or less. Among the sub-studies, 33 basic studies did not value housing. From this it can be concluded that it can be understood that the quarters were loved by numerous undergraduates and in good condition.

Table 4.13 records the evaluation of the exercises conducted during the camp and now has a significant underestimation. According to 184 sub-studies, the assessment level is 10, 70 sub-studies were reviewed with 9, 43 sub-studies were checked with 8 points and 7 or less were evaluated with 27 sub-studies. Examination of the sub-studies shows that the exercises were dynamic and rewarding.

Table 4.14 stores estimates of the nature of mass-mediated events in sub-studies. You can see that 183,325 sub-studies rate quality by limiting 10. This

means that more than half of the sub-studies rate the cases as “incredible”. The remaining points are given below: 49 sub-studies 9, 36 sub-studies 8, 22 sub-studies 7 points and 23 sub-studies 6 points or less. As these data show, may well conclude that joukkostimulaatiotilaisuudet formed well, and that they are justified by concerns over the alaopinnoista.

The age of undergraduate study is almost the same, so we did not study age. We had no information on male / female characterization, so we did not specify the proportion.

Following the above study, the goals of the roadmap have been determined and can be extrapolated to other teaching camps around the world.

## 4.6 Recommendations

We recommend improving the "EduCon" framework by:

- Adds attention to materials science and arithmetic, as they are mostly selected urban areas.
- In terms of nutrition, everything is great, and there is no compelling reason to improve it far better (just keep reliability, don't make it worse)
- More attention will be paid to training in collecting eyeglasses and football. These are the most ideal exercises for members. Improve the Nauryz Gathering collection better, given that a few members chose it as a horrific activity.
- Make arithmetic and materials science exercises even more interesting, given that the level of individuals who have chosen these topics as malignant ones indicates a risky question. In any case, make it the same enthusiasm as in the history exercises.
- Everything is great about the residence permit. Be that as it may, the company incorporates reflections and WIFI Backing to make it much better.

# 5. The second study

## 5.1 Objectives

To anticipate input across members concerning the nature of educating in "EduCon" Camp, as indicated by close to home data, most loved exercises, and consequences of tests that have been led for 4 months.

## 5.2 Limitations

target members were understudies from Kazakhstan, age run between 10-18 years of age, who took an interest in the overnight-remain camp "EduCon" during spring 2018. The investigation apparatuses contained an imaginative outdoors game plan program, and a poll in a paper-based structure. The poll comprised of inquiries concerning social and learning fulfillment as their sentiment about the residence, top choice/unfavourite subject, food, movement, occasion, and the other related ones. In the wake of gathering the information, it was moved to "Google Sheet" physically. The information was utilized as a gracefully to the AI part, where the choice trees model has been used.

## 5.3 Justification of the picked methodology

The AI calculation, which has been used during the expectation part, is "choice trees". It has been chosen because of simplicity of cognizance from a human point of view providing created rules of expectation on its premise.

Choice Trees is an AI calculation, which works by organizing a tree, where the root contrasts to the segment and the most amazing information increment, and leaves identify with the goal worth, however the route from the leaf till the root

can be considered as requests that the model posture in order to foresee target variable for some gave data.

We used a combination of choice tree models in regards to changing its parameters to plan and test the models to anticipate the information assessment of an understudy about the camp indicating the progressive part. It moves between 0 to 5.

We partitioned the information into preparing and testing part, where 67 % of dataset have been assigned to preparing part, though remaining 33 % to the testing stage.

## 5.4 Used technologies

To accomplish the above-mentioned purposes Python and Scikit Learn library have been utilized.

Since artificial intelligence was previously recognized as an order in the mid-1950s, AI has become a major research area. There are two reasons for this noticeability. Learning ability is a sign of insightful behavior, so any desire to understand insight as a miracle should include an understanding of learning. Moreover, training provides a potential approach to creating high-performance structures.

Learning research consists of various subfields. In one unusual, there are universal structures that show their own exhibition and strive to improve it by changing the parameters of the interior. This methodology, usual for a huge degree of early academic work, provided self-improvement projects for fussing, adjusting posts, solving problems, in addition, numerous different spaces. A very unique methodology involves finding out how you can get organized information in the form of ideas - separation networks or creation rules.

The functional significance of the AI of this latter species was emphasized by the emergence of information-based core structures. As their name implies, these structures are driven by information that is explicit, not understood in the calculations. The information, which is expected to serve as the basis for the creation of the basic structures, has been classified through enhanced communication between the space expert and the information engineer. While the standard rate of information clarification using this strategy is a couple of rules per person per day,

a specialized structure for a stunning enterprise may require hundreds or even a lot of such recommendations. Obviously, the method of meeting with obtaining information cannot correspond to the interests of the masters; Feigenbaum calls this the bottleneck issue. This observation enlivened the study of AI methods as methods of information clarification. Carbonell, Michalsky and Mitchell identify three main dimensions by which the framework of AI can be characterized:

- the hidden learning systems utilized;
- the portrayal of information obtained by the framework;
- the application area of the framework.

Accepting these key points ago the request, the scope of these structures is not limited to a specific area of scientific action, such as chemistry or chess; they can be applied to any such zone. Although they are accordingly widely useful environments, all the applications they access include a feature. The result of the training is a small part of the procedural information, which can refer an inconspicuous hitherto inconspicuous article to one of a predetermined number of disjoint classes. Examples of grouping of assignments are: determination of an ailment by side effects, when classes can be either different states of the ailment, or potential treatment; Solving a game-theoretic evaluation of a chess position with classes won for white, lost for white and draw; and The choice from barometric perception is whether an extreme storm is unbelievable, conceivable or probable.

This may give the impression that characterization instructions are only a small subset of procedural tasks, but even exercises such as organizing a robot can be overestimated as problems with organization.

People from this family are strongly portrayed because they portray the information they purchase as trees of choice. This is usually a simple informational formalism that does not have the expressive intensity of semantic systems or other images of the first request. As a result of this straightforwardness, the teaching philosophies used in The TDIDT family is significantly less complex than those used within the framework that can express the consequences of their training in an even more impressive language. By the way, it is still possible to present information as trees of choice, which are ready to solve difficult issues of great importance.

The hidden system is gradually benefiting from models. The frameworks are given many cases related to the grouping problem, and a selection tree is created, starting from the top, guided by the repeatability data in the models, but not by the specific query in which the models are provided. This looks different with respect to gradual methods, for example, those used by MARVIN, in which the exchange with the teacher for "partial study" of the right ideas is continued, and used by Winston, in which the models are cut apart in turn, each introduces a slight change in the idea of creation; in both of these structures, the query in which the models are presented is usually of great importance. Thus, the structures described here for the study of structures in these models should be able to verify and rethink each of them at numerous stages of training. Other notable projects that share this information-based methodology include BACON.

In a brief overview, at this point, the wireframes shown here create selection trees for grouping orders. These trees begin at the base of the tree and continue to its leaves. The palindromic family name emphasizes that its individuals perform top-down induction of decision trees.

The model articles from which the grouping rule is created are known as evaluations of many properties or attributes, and selection trees are thus passed up to these equivalent attributes. The models themselves may be accumulated in two different ways. They can come from the current database, which forms the history of perceptions, for example, understanding records in a certain treatment area that have gathered in the analytic community. Objects of this kind give a reliable factual picture at the same time, since they are not compiled at all, they can be repeated or throw away remarkable cases that were not observed during the recording. On the other hand, the elements can be a painstakingly spaced arrangement of training exercises models organized by the master of the field, each of which has a certain significance for the general and correct grouping rule. A wizard can go to great lengths to maintain a strategic distance from excess to include cases of unusual cases. Although a group of structures will manage assortments of any kind in an acceptable manner, it should be mentioned that before

TDIDT structures were planned with a "verifiable record" approach as a top priority, however, all the structures presented here are currently regularly used with training sets of exercises.

CLS creates a selection tree that attempts to limit the cost of grouping items. These costs have two types of segments: the cost of evaluating the decision to evaluate property A shown by the element, and the cost of misclassifying the choice of whether an article has a place with class J when its genuine class is K. CLS uses a forecasting methodology. like minimax. At each stage, the CLS explores the space of possible selection trees with a fixed depth, selects an action to limit costs in that limited space, and at that moment moves one level down in the tree. Depending on the depth of choice, CLS may require a significant calculation measure, however, it was possible to find unobtrusive examples in the articles that appeared.

ID3 is one of the progression of projects created from CLS in reaction to the difficult task of adoption presented by Donald Michie, namely. choose one of the design-based highlights, whether the specific chess position in the King Rook vs. Lord Knight endgame is lost for the Knight side with a fixed number of handles. The complete ID3 image is shown in Section 4, so it's enough to note here that it inserts a tree-building strategy into an iterative external shell and refuses a cost-based approach to CLS with data-driven evaluation work. The tendons of the upper leg are an ID3 spec. CLS and ID3 require that each property used to represent objects has only qualities from a predefined set. Regardless of the properties of this type, ACLS licenses properties that have the unhindered qualities of an integer. The ability to manage properties of this kind has allowed ACLS to be applied to complex tasks, such as image validation.

The partner also recognizes ID3 as its immediate predecessor. It varies from ID3 from different points of view, some of which are discussed in detail in subsequent segments. In addition, Associate summarizes all the advantages of ACLS, evaluating them, providing properties with constant (genuine) values. Unlike requiring classes to be disjoint, ASSISTANT allows them to create a chain of importance with the goal that one class can be the best separation of the other. The right hand does not form the selection tree iteratively, as in ID3, but includes calculations for choosing a "decent" training set from the available articles. The right hand has been used in several clinical settings with promising results.

Most of the three frameworks in the figure are based on ACLS subsidiaries. Although they do not fully advance the basic hypothesis, they combine numerous understandable designs and utilities that facilitate the creation and use of selection

trees. Everyone has surprisingly modern victories. For example, the Westinghouse Electric water reactor division specializes in fuel marketing applications in which the organization was able to maintain revenue "exceeding ten million dollars a year" using one of them.

One of the ways to cope with the task of listing above is to create all conceivable choice trees that accurately characterize the training set and choose the least difficult of them. The number of such trees is limited, but enormous, so this methodology would simply be achievable for small reception enterprises. ID3 was intended for the opposite end of the range, where there are many characteristics, and the training set contains many elements, however, where a reasonably large selection tree is required, there is not a lot of calculation. For the most part, it was discovered that he was building the base trees of choice, but the methodology he used could not guarantee that the best trees were not ignored.

The basic structure of ID3 is iterative. A subset of the training set, called a window, is selected aimlessly, and a selection tree is created from it; This tree effectively groups all the articles in a window. All other articles in the training set at this point are grouped using the tree. If there is a possibility that the tree will offer the correct answer for each of these articles, then at this point it is suitable for the whole set of preparations, and the procedure ends. If not, the definition of inaccurately grouped articles is added to the window, and the procedure continues.

Along these lines, the right selection trees were found after only a few cycles to prepare sets of up to thirty thousand items, depicted up to 50 characters. Accurate evidence recommends that the tree of the right choice, as a rule, be detected with this iterative technique much faster than by lawfully forming the tree from the entire preparatory set.

However, O'Keefe noted that the iterative system cannot be combined in the last tree, unless the window can be developed to include the entire preparatory set. This potential limitation has not yet arisen gradually. The test solution is urgent if the selection tree should be the base. On this occasion, the test will be limited to expanding the character score, so choosing a test reduces to choosing a characteristic for the tree base. In the primary acceptance programs, when applying for identity cards, the "seat for pants" assessment was used, which worked reasonably. Following Peter Gax's suggestion, ID3 used information technology

based on two suspicions. Let C contain p objects of class P, and moreover, n of class N. An exceptional case arises if C does not contain elements with some specific value Aj of A, which gives the emptiness of Cj. ID3 calls such a sheet "invalid" for the purpose that it ignores the grouping of any object found on this sheet. Superb layout summarizes from the set C from which Cj comes from and makes this sheet more continuous class in C.

The heuristic value of the ID3 trait selection can be estimated by the ease of subsequent selection trees or, more specifically, by how well these trees express the true relationships between class and characteristics, as shown by the accuracy with the protests that they order, except for those in the preparatory set (their prediction of accuracy). A clear technique for studying this predicted accuracy is to use some part of a given arrangement of objects as a preparatory set and check the subsequent selection tree for the rest.

Several analyzes of this kind have been made. In one region, 1.4 million chess positions depicted in 49 honorary diplomas offer 715 separate articles divided by 65 percent: 35 percent between classes. This area is moderately amazing, as the tree of the right choice for each of 715 articles contains about 150 nodes. In preparing the kits containing 20 percent of these 715 items, they were selected indiscriminately, they created selection trees that effectively characterized more than 84 percent of inconspicuous items. In another form of similar space, 39 characteristics gave 551 unmistakable goods with the right choice of wood of comparative size; the preparation of sets of 20 percent of these 551 articles made it possible to select trees of virtually indistinguishable accuracy. In a simpler space (1987 objects with the right tree selection of 48 knots), I accidentally chose cooking sets containing 20  7 . Of these objects, trees were selected that effectively characterized 98 percent of the hidden objects. In each of the three cases, decision trees accurately reflect the valuable (rather than irregular) relationships present in the information.

## 5.5 Results

In our first endeavor, table 5.1 shows outcomes we accomplished by grouping report:

We chose to reject 0 esteem because very one gave 0 for the camp, however it

Table 5.1: The first experiment

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 3 | 1.00 | 0.25 | 0.40 | 4 |
| 4 | 0.28 | 0.20 | 0.23 | 35 |
| 5 | 0.82 | 0.89 | 0.85 | 155 |

Table 5.2: The second experiment

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 3 | 1.00 | 0.25 | 0.40 | 4 |
| 4 | 0.22 | 0.23 | 0.22 | 35 |
| 5 | 0.82 | 0.83 | 0.85 | 155 |

was simply a substitution variable for nulls. There was likewise no 1 and 2 grades, with the goal that we didn't consider them in the order report.

To manage the highlights, which esteems have been spoken to in a printed way, we utilized pandas sham factors.

From that point forward, we chose to begin the hyperparameter enhancement stage. The primary parameter which we enhanced is the most extreme profundity of our decision tree model. Profundity term alludes to the number of edges from the root hub (which was split as indicated by a strategy of choosing the element with the most elevated data gain) till the leaf, and table 5.2 shows the outcomes which we accomplished for this situation

In examination with the past outcomes, we can see that for the class 3 results did not change, because the quantity of individuals who cast a ballot 3 for the association of concentrate in the camp is exceptionally low in correlation with classes 4 and 5. As to class 4, we can see that exactness diminished a tad, review and f1 score expanded a tad. As to class 5, just review diminished a tad. It demonstrates that for this situation advancing the most extreme profundity of the tree isn't basic since it didn't prompt some extraordinary changes.

With regards to the best incentive for most extreme profundity, we found that it is 7 utilizing a network search strategy.

The following intriguing hyperparameter which we needed to upgrade was the base number of parts, which alludes to the base number of tests of information which are expected to part an inner hub of the decision tree model.

Table 5.3 shows the outcomes, which we accomplished :

Table 5.3: The third experiment

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 3 | 0.00 | 0.00 | 0.00 | 4 |
| 4 | 0.25 | 0.03 | 0.05 | 35 |
| 5 | 0.80 | 0.98 | 0.88 | 155 |

Table 5.4: The fourth experiment

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 3 | 0.00 | 0.00 | 0.00 | 4 |
| 4 | 0.00 | 0.00 | 0.00 | 35 |
| 5 | 0.79 | 0.99 | 0.88 | 155 |

We can see that review of class 5 have been definitely expanded from 0.83 to 0.98, as to class 3 and 4 these are of less significance, in light of the fact that the explicitness of the dataset is that practically all members gave 5 vote, which lessens the significance of different evaluations because of their not many numbers in the information.

We found that 192 is the most ideal number of parts for this dataset.

Presently, we thought imagine a scenario where we perform network search on the crossing point of most extreme profundity and least examples for parting. Table 5.4 shows the outcomes we accomplished :

# 6. The third study

## 6.1 Methods and materials

### 6.1.1 Dataset description

Dataset comprises of 645 examples. Each example speaks to individual data and aftereffects of tests of a camp member.

### 6.1.2 Used technologies

- Google sheets
- Advanced google sheet capacities (vlookup, importrange)
  - vlookup: Vertical question. Searches down the essential segment of a range for a key and returns the estimation of a predefined cell in the section found.

### 6.1.3 Inner mechanics of the system

Figure 7.1 shows how the information assortment framework functions at an elevated level. In the first place, the information is passed to the vlookup work as a parameter, the capacity at that point makes an information recovery framework and gives a google sheets representation instrument to demonstrate a histogram of the item to be attracted the test aftereffect of the individual partaking in the camp.

### 6.1.4 Justification of the picked methodology

#### Rationale of picking google sheets

Google Sheets makes boarding simple for clients. You don't need to pay for classes or even contribute vitality by taking transversal advisers for investigate Google Sheets. It has an intuitive interface that makes your life a lot more straightforward.

Google Sheets stands apart most unmistakably from Excel on account of its incredibly planned exertion. Various people can manage a comparable twofold tallying spreadsheet without releasing or befuddling. You can even access a spreadsheet from your record on various gadgets simultaneously without interference. You can likewise visit clients since you've currently snapped a photo on a spreadsheet with you, which makes correspondence significantly more skilled than a progression of messages.

Also, Google Sheets is facilitated with all other Google applications. This implies you can exploit applications like Google Translate or Google Finance inside Google Sheets with fundamental guidelines.

Not in the slightest degree like Microsoft Excel, you have made a fix history in Google Sheets that permits you to keep tabs on your development.

You can helpfully get to your spreadsheets through Google Drive, regardless of what PC you're on. This disposes of the need to send records to and fro through email.

In synopsis, there are numerous different things in Google Sheets, so on the off chance that you get an opportunity to discover something about its most basic customization defect, there's likely an extra that you can download to stress over it.

#### Rationale of imagining data utilizing histograms

- Histogram makes our task less difficult to recognize different data, the repeat of the data occurring in the dataset and classes which are difficult to

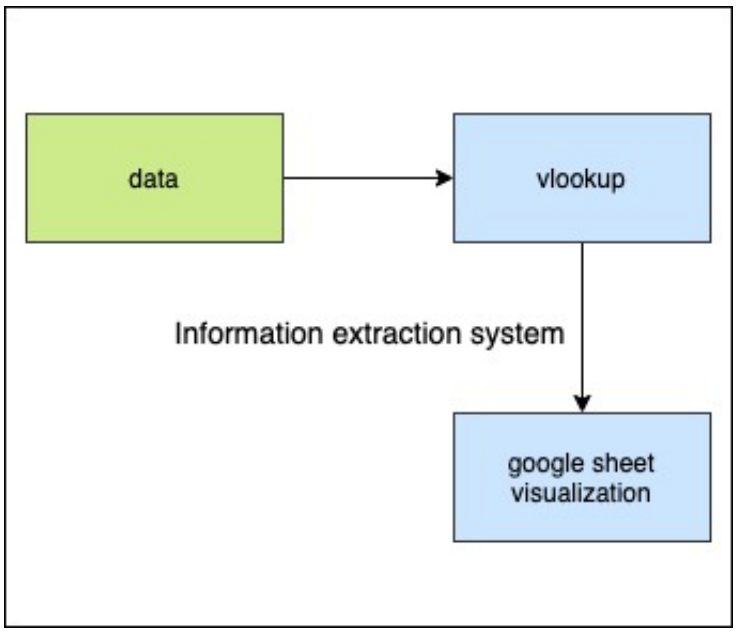


Figure 6.1: High level dynamic of the system

interpret in an inconceivable structure.

- It envisions the allotment of the data.
- When we have enormous datasets it might be viably imagined using a histogram.
- It in like manner gets an understand of the skewness of the data.

### 6.1.5 Limitations

- Unable to adjust to structure of new dataset without inclusion of engineers

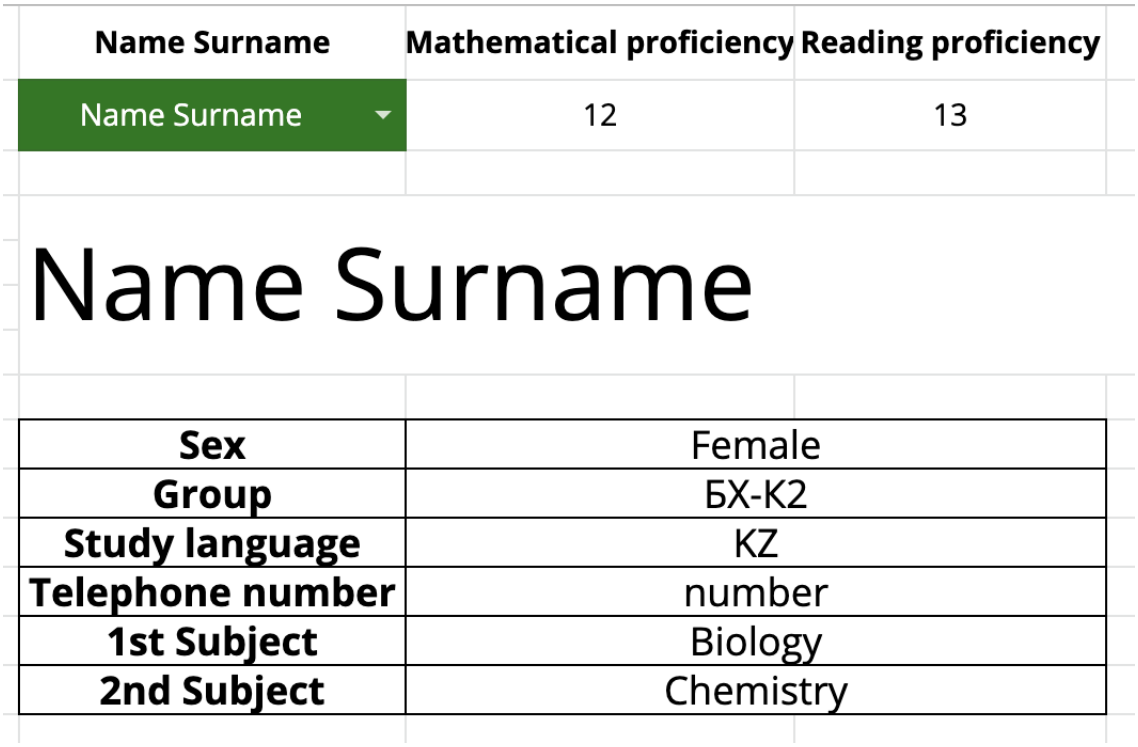


Figure 6.2: Extraction of personal information

## 6.2 Results and discussion

Figure 6.2 shows that the system can choose any person from the dataset and automatically show in an easy-to-read way personal information of a camp participant.
Figure 6.3 shows a distribution of grade for exams according to subjects of a camp participant.
Figure 6.4 shows a tool for displaying average exam results across all camp participants by month for each city in the dataset.

The possible implications of the findings might result in simplification and improving the management process of educational camps worldwide without installing special software or packages just by opening a google sheet link and getting access to a user-friendly interface to get information about each camp participant.

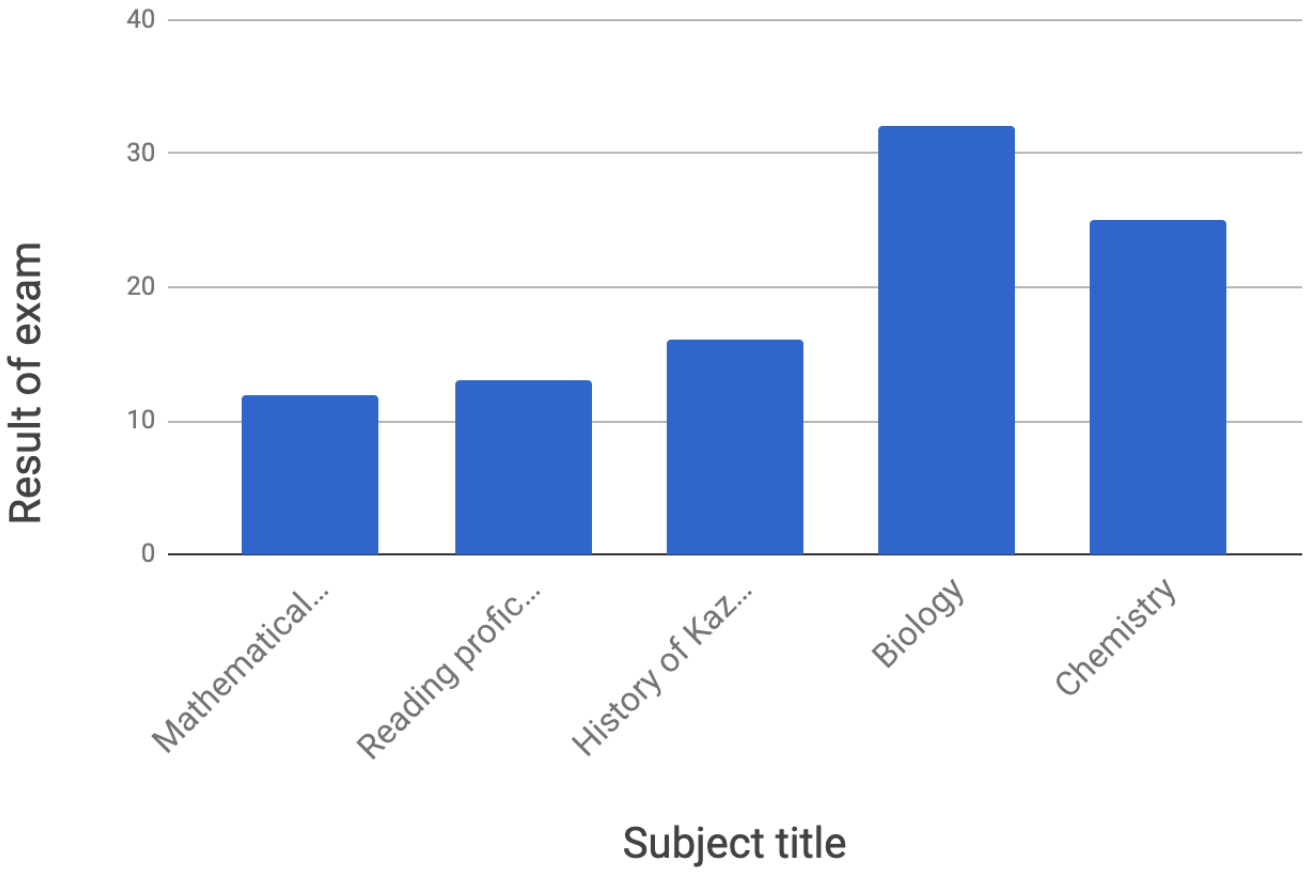


Figure 6.3: Histogram of relationship of exam results to a subject title

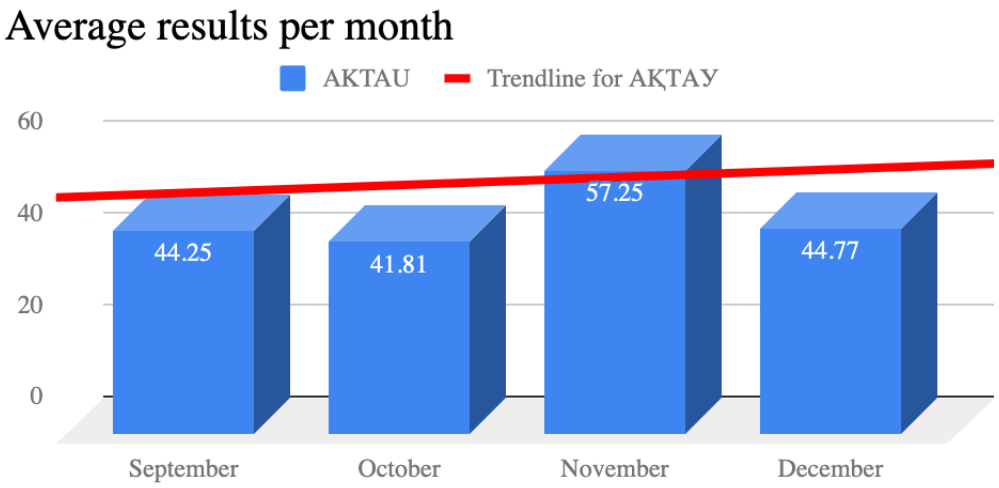


Figure 6.4: Average monthly exam results for each cities

# 7. The fourth study

## 7.1 Overview

In this work, there were dissected overviews of the Paragon IT hackathon. In light of that, there were acquired proposals with regards to how to improve this hackathon significantly further. The discovered outcomes can be extrapolated into different hackathons, particularly in Asia. Likewise, there was done insights of members as their sexual orientation, age, level of met desires, and the probability of going to another hackathon in the wake of experiencing this one. At long last, the fundamental difficulties and things they learned were composed with their essential inspiration to take an interest.

Hackathons are very popular all around the globe [16]–[18]. They are uncommon exercises, which ordinarily last around 24 hours, where groups are framed to introduce their imaginative answers for the given issues in the IT space. In this paper, there was analyzed the poll aftereffects of the Paragon Hackathon, which was carried on January 10-11, 2020 at Paragon International University.

ParagonIU Hackathon was a 2-day coding run that provoked members to build up a web application by the utilization of Facebook API.

The inquiry which the work pose is the thing that authoritative changes ought to happen to build up the hackathon significantly further. The got outcomes can possibly be extrapolated into different hackathons, especially in Asia, because of the higher probability of the similitude of members.

Paragon International University (once in the past known as Zaman University) was developed in 2010, as another person from Zaman informational associations (by and by under Paragon Education Company). Zaman University was rebranded and was officially seen as Paragon International University by a sub-presentation of the Royal Government of Cambodia on 28 January 2019.

With a guarantee to significance, Paragon International University tries to be the principle school in Cambodia to the extent demonstrating quality, research and system organizations, to give all around apparent undertakings and produce a wonderful pool of HR that will add to the general progression and improvement of Cambodia and its occupants. In seeing an overall world and the prerequisite for overall correspondence, since its start, the language of direction at Paragon International University has been English.

## 7.2 Methods and materials

### 7.2.1 Dataset description

The dataset was about the review aftereffects of 31 hackathon members. The study was isolated into 2 sections. The initial segment concerned one bit of members, though the other part concerned the staying one. The inquiries were about sex, age, level of met desires, the probability of going to another hackathon in the wake of experiencing this one, what was the greatest test for them, did they get new techniques and information during the hackathon, what did they learn new about the Facebook engineer, and any remarks or recommendations.

### 7.2.2 Used technologies

To make an analysis, there was used Python programming language, Pandas library.

- Python is a deciphered, elevated level, all around valuable programming language. Made by Guido van Rossum and first released in 1991, Python's structure thinking underscores code clarity with its striking usage of basic whitespace. Its language creates and object-orchestrated procedure expects to help programming engineers with making clear, savvy code for close to nothing and enormous extension adventures

- In PC programming, Pandas is an item library made out of the Python programming language for data control and assessment. In particular, it offers data structures and exercises for controlling numerical tables and time course of action. It is free programming released under the three-stipulation BSD permit. The name is gotten from the articulation "load up data", an econometrics term for instructive assortments that fuse discernments over various timespans for comparative individuals.

### 7.2.3 Justification of the chosen methodology

Python is viewed as a standout amongst other programming dialects for logical purposes. It is anything but difficult to compose a content on it and execute it.

Python can be effectively propelled on various stages as Google Colab, Jupyter Notebook, Deepnote, and so on.

Pandas is a Python library for information wrangling and investigation. Loads of information investigation capacities are pressed into discrete modules on it, which makes the way toward composing information examination code smooth.

## 7.3 Results and discussion

Table 7.1 shows average age, level of met expectations, and likelihood of attending another hackathon collected from the participants.

Table 7.1: Average results for the first survey

| | |
|---|---|
| Age | 19.400 |
| Were your expectations met? | 3.875 |
| How likely do you attend to another hackathon? | 3.875 |

Table 7.2 shows the same results but for the second part of the survey collected from the other subset of participants.

Table 7.2: Average results for the second part of the survey

| | |
|---|---|
| Age | 19.500 |
| How likely do you attend to another hackathon? | 3.800 |

For each of the parts of the survey, there was analyzed the data by gender. Table 7.3 shows average results for the first part of the survey for only male

Table 7.3: Average results for the first part of the survey for male

| | |
|---|---|
| Age | 19.410 |
| What were your expectations? | 3.923 |
| How likely do you attend to another hackathon? | 4.076 |

Table 7.4 shows average results for the first part of the survey for only female

The results for table 7.3 and table 7.4 did not differ so much.

One of the respondents decided not to share his/her gender. Conclusion based on one record might not be representative, therefore it was decided not to include it to this section.

### 7.3.1 Exploratory data analysis

#### Categorical columns exploration

In the first part of the first survey, there were 16 rows and 43.75 percentage of missing values. As to the second part of the first survey there were 15 rows and 6.66 percentage of missing values.

Table 7.4: Average results for the first part of the survey for male

| | |
|---|---|
| Age | 19.000 |
| What were your expectations? | 4.000 |
| How likely do you attend to another hackathon? | 4.000 |

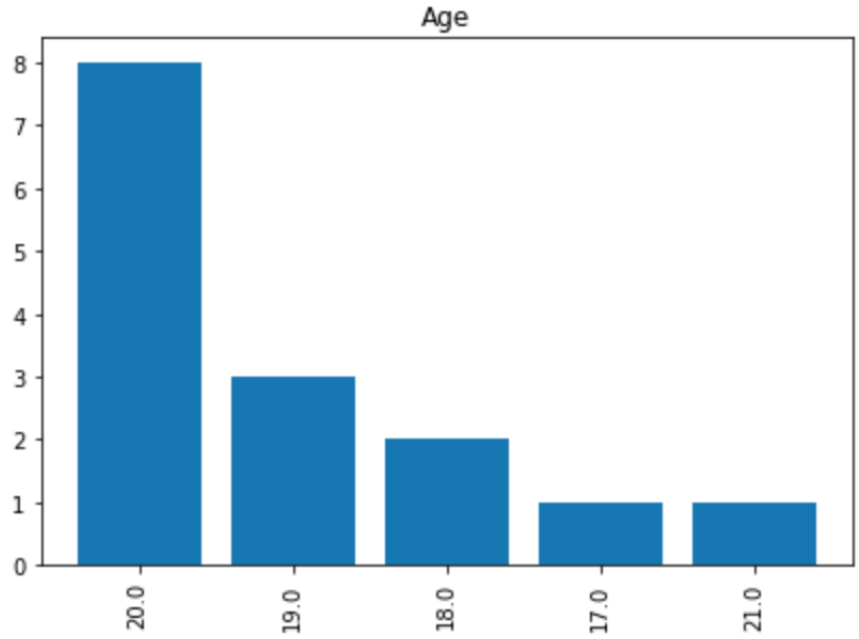


Figure 7.1: Distribution of age for the first part of the first survey

Figure 7.1 shows distribution of age for the first part of the first survey, whereas figure 7.2 shows distribution of age for the second part of the first survey. It can be observed, that most of participants were in 20, and the least of them were in 17.

Figure 7.3 shows distribution of gender for the second part of the first survey. It is shown that most of participants are males.

Figure 7.5 shows distribution of whether this is the first hackathon for a participant for the second part of the first survey, whereas figure 7.4 shows distribution of whether a participant heard about Facebook Developer API before for the same data. These questions were not asked in the first part of the survey, therefore they were not added to the analysis. For most of participants this was the first hackathon, whereas whether they have heard about Facebook API developer is that almost half of participants did not know about it, and the remaining half knew.

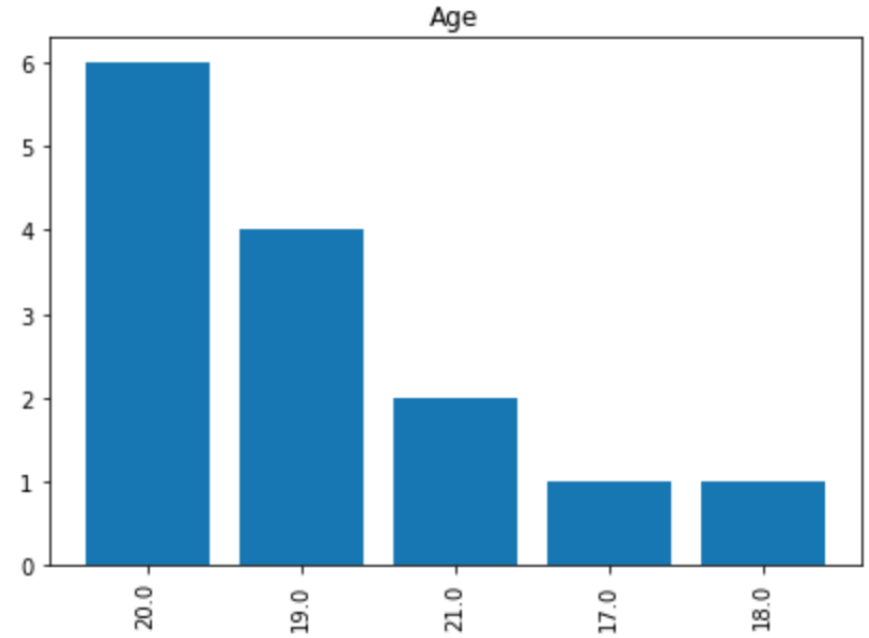


Figure 7.2: Distribution of age for the second part of the first survey

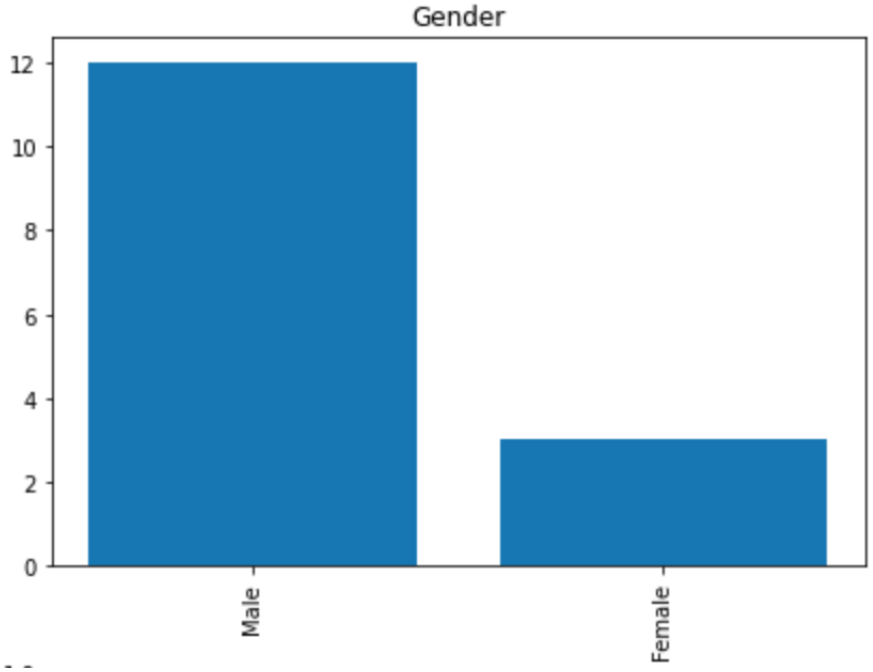


Figure 7.3: Distribution of gender for the second part of the first survey

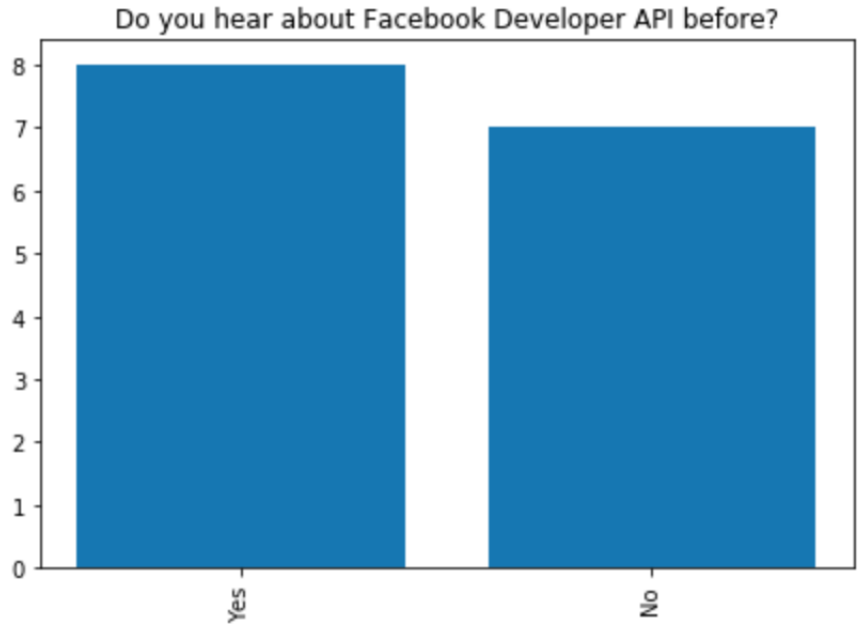


Figure 7.4: Distribution of whether a participant heard about Facebook Developer API before for the second part of the first survey

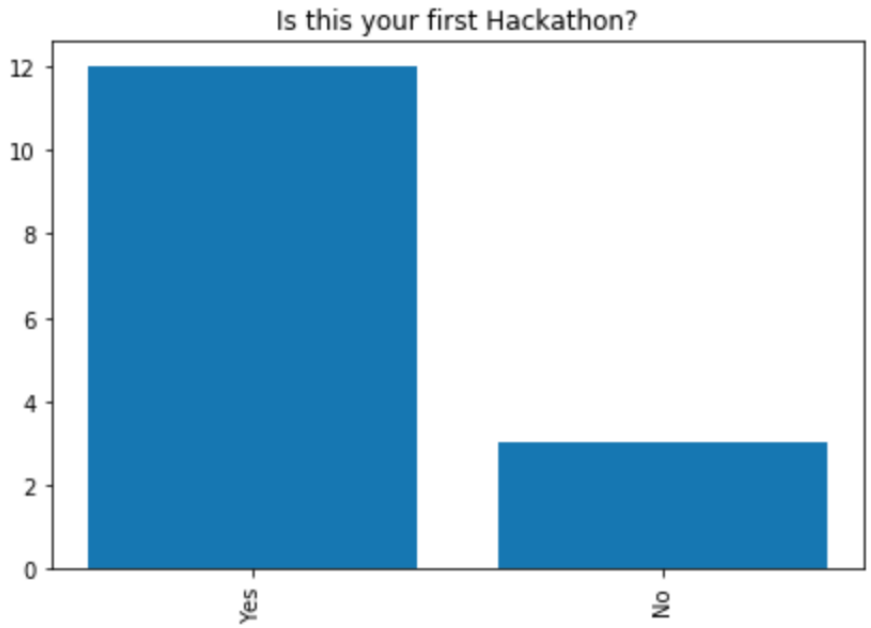


Figure 7.5: Distribution of whether this is the first hackathon for a participant for the second part of the first survey

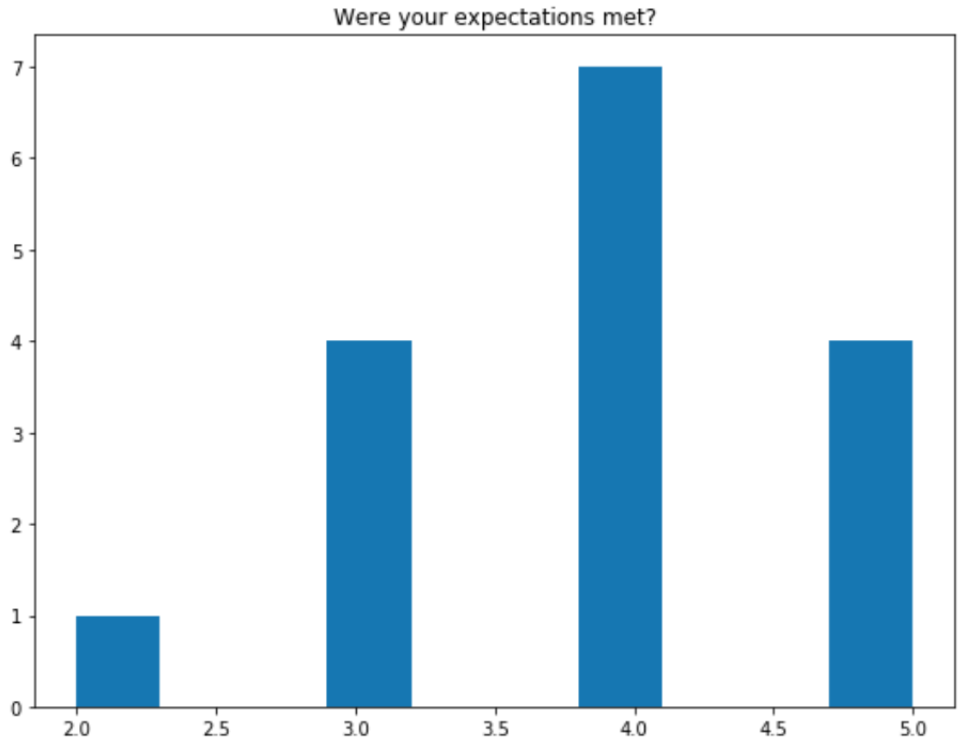


Figure 7.6: Distribution of level of met expectations for the first part of the first survey

### Numerical columns exploration

Figure 7.6 shows distribution of level of met expectations for the first part of the first survey, whereas figure 7.7 shows distribution of likelihood of attending another hackathon after attending this one for the same data.

figure 7.8 shows distribution of likelihood of attending another hackathon after attending this one for the second part of the first survey. Distribution of level of met expectations for the second part of the first survey was not collected, therefore its visualization not added to this analysis.

### Looking at differed data

Figure 7.9 shows dissemination of sexual orientation as far as level of met desires for the initial segment of the principal review.

Figure 7.10 shows appropriation of sex as far as probability of going to another hackathon subsequent to taking an interest in this one for the second piece of the primary overview.

By and large, it tends to be presumed that the members were content with the association of the hackathon. Be that as it may, what makes it not impeccable

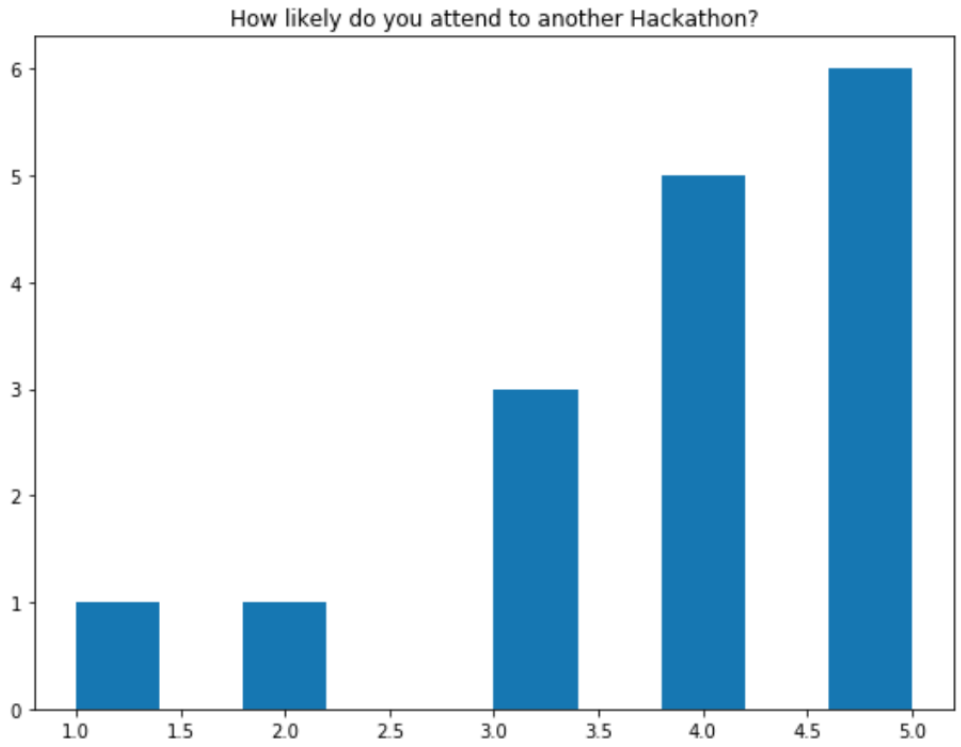


Figure 7.7: Distribution of likelihood of attending another hackathon after attending this one for the first part of the first survey

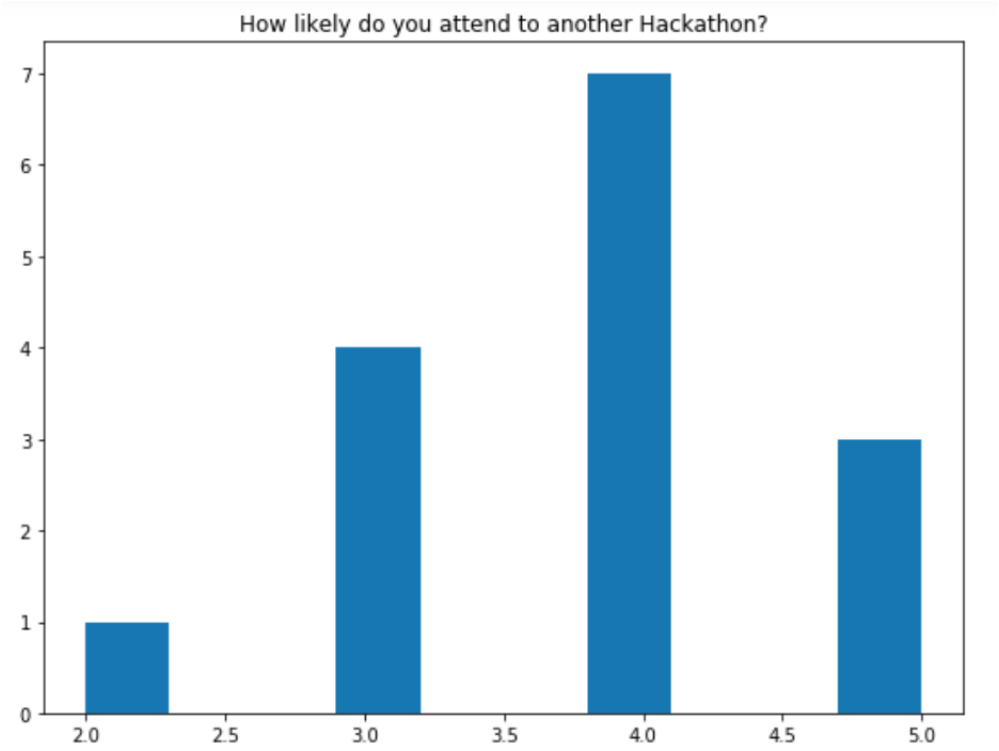


Figure 7.8: Distribution of likelihood of attending another hackathon after attending this one for the second part of the first survey

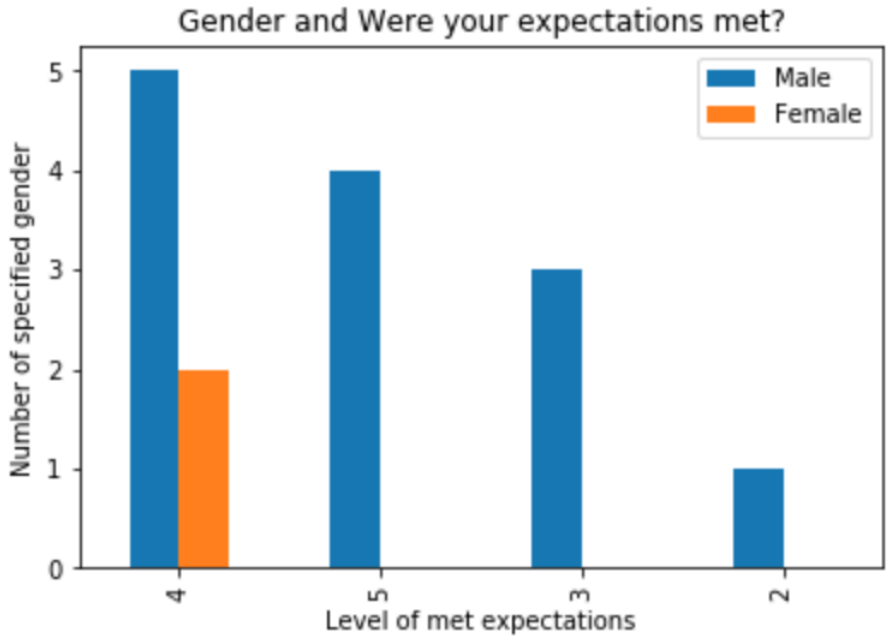


Figure 7.9: Distribution of sexual orientation and level of met desires for the initial segment of the first survey

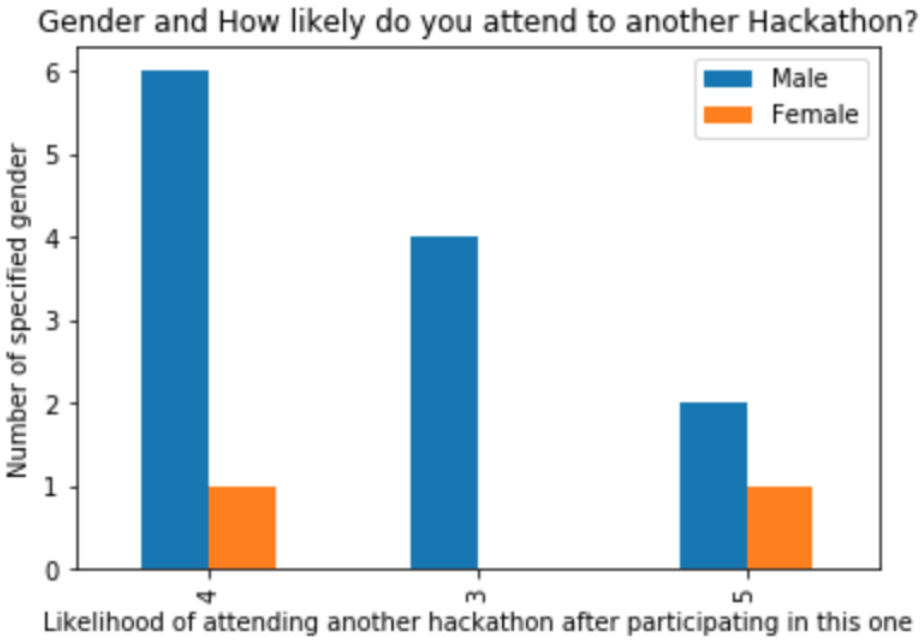


Figure 7.10: Distribution of sexual orientation and level of met desires for the initial segment of the first survey

may be the accompanying focuses which depend on the remarks and proposals composed by the members:

- The hackathon was not centered more around the data gave to understudies. They didn't know precisely what to do and what the Hackathon anticipates from them
- Any group ought not have a great deal of members. 2-3 man for every group is by all accounts enough
- Number of members ought to get expanded
- Number of classifications ought to get expanded

Handiness of hackathon from the point of view of members is affirmed similarly as in [16]–[18], [22]–[27], anyway new bits of knowledge have been acquired based for this particular hackathon.

Regarding the principle challenges that members experienced are :

- Understanding the business idea to use in an undertaking
- Many things to gain proficiency with at the same time
- Whether to concentrate on innovation or business in an undertaking
- Time limit
- Lack of coding abilities
- Distractions

With respect to the things that the members got the hang of as indicated by the review are:

- Construction of Chatbot
- Spark AR
- Messenger AI bot
- Facebook API utilization and component

- How to fabricate and transfer game utilizing SDK, execution, and preparing of chatbot

As far as the essential inspirations to take an interest in a hackathon are:

- New experience (the vast majority of the members composed this was their essential inspiration)
- Friends urged to join in
- To be a piece of a social gathering
- Work on true issues

Organizing hackathons is a cumbersome issue. In this work, there was coordinated Paragon IT hackathon. Most by far of the outline results were sure. Regardless, some definitive changes must happen to make it by a wide margin predominant as augmentation of individuals, classes; decrease of individual per gathering; giving point by point information about hackathon.

# 8. Conclusion

Implementation of id3 algorithm in the educational organizations domain in the efficient way is a cumbersome problem. The experimental results showed that the most optimal maximum depth for the decision tree is 3 and the most optimal minimum number of splits is 192. Based on that, there was achieved optimum results of precision, recall, and f1 score machine learning metrics that vary between 75 to 98 depending on the change of tunable variables of the ID3 algorithm. Also, information extraction system was built and recommendations to enhance the educational organization based on the computational data analysis were successfully built.

# References


[1] A. Kney, J. Tatu, M. marlin, and X. Meng, *Transforming stem to steam (work in progress): How a traditionally run stem camp succesfully incorporated the arts into its framework*, Integrated STEM Education Conference (ISEC), 2016.

[2] J. Gregg and T. Chen, *Peer: Enriching secondary engineering education through a college summer camp*, nternational Conference on Microelectronic Systems Education (MSE'05), 2005.

[3] J. M. J.C. Sener and L. Aburusa, *Process-education-based teaching and learning at the idaho engineering science camp*, 32nd Annual Frontiers in Education, 2002.

[4] *Google trends*, Available: https://trends.google.com/trends/, 2018.

[5] L. Zhan and X. Jiang, *Survey on event extraction technology in information extraction research area*, 2019.

[6] C. Feilmayr, *Text mining-supported information extraction: An extended methodology for developing information extraction systems*, 2011.

[7] X. Zhang, Y. Wang, and L. Wu, *Research on cross language text keyword extraction based on information entropy and textrank*, 2019.

[8] Q. Zhu and X. Cheng, *The opportunities and challenges of information extraction*, 2008.

[9] L. Liu, G. He, X. Shi, and H. Song, *Metadata extraction based on mutual information in digital libraries*, 2007.

[10] A. Serek, M. Zhaparov, and S.-M. Yoo, *Analysis of data to improve system of an educational organization*, 2018.

[11] C. Demetry and D. Nicoletti, *Camp reach: An engineering summer camp for middle school girls*, 1997.

[12] J. Gregg and T. Chen, *Peer: Enriching secondary engineering education through a college summer camp*, 2005.

[13] L. A. Lyon and E. Green, *Women's attraction to coding boot camps*, 2018.

[14] Y. Gamo, M. Yano, and K. Sakai, *Mitsubishi kougaku-juku: A technology training camp for engineers of the future*, 1991.

[15] *Educon*. [Online]. Available: `https://educon.kz/`.

[16] A. Decker, K. Eiselt, and K. Voll, *Understanding and improving the culture of hackathons: Think global hack local*, 2015.

[17] M. Lin, M. Otten, D. Pleiter, J. E. Stone, J. Lucio-Vega, M. Zingale, and F. Foertter, *Best practices in running collaborative gpu hackathons: Advancing scientific applications with a sustained impact*, 2018.

[18] S. Saravi, D. Joannou, R. S. Kalawsky, M. R. N. King, I. Marr, M. Hall, P. C. J. Wright, R. Ravindranath, and A. Hill, *A systems engineering hackathon – a methodology involving multiple stakeholders to progress conceptual design of a complex engineered product*, 2018.

[19] A. T. R. ; A. L. Miller, *A collaborative approach to offering summer engineering camps for middle school students*, 2009.

[20] A. Harriger, A. J. Magana, and R. Lovan, *Identifying the impact of the spirit program in student knowledge, attitudes, and perceptions toward computing careers*, 2012.

[21] L. A. Lyon and E. Green, *Women's attraction to coding boot camps*, 2018.

[22] A. Decker, K. Eiselt, and K. Voll, *Understanding and improving the culture of hackathons: Think global hack local,* 2015.

[23] J. Tandon, R. Akhavian, M. Gumina, and N. Pakpour, *Csu east bay hack day: A university hackathon to combat malaria and zika with drones,* 2017.

[24] S. C. et al., *Best practices in running collaborative gpu hackathons: Advancing scientific applications with a sustained impact,* 2018.

[25] M. Raatikainen, M. Komssi, V. d. Bianco, K. Kindstöm, and J. Järvinen, *Industrial experiences of organizing a hackathon to assess a device-centric cloud ecosystem,* 2013.

[26] K. Gama, B. Alencar, F. Calegario, A. Neves, and P. Alessio, *A hackathon methodology for undergraduate course projects*, 2018.

[27] S. Saravi, D. Joannou, R. S. Kalawsky, M. R. N. King, I. Marr, M. Hall, P. C. J. Wright, R. Ravindranath, and A. Hill, *A systems engineering hackathon – a methodology involving multiple stakeholders to progress conceptual design of a complex engineered product*, 2018.