**[](http://www.akademik.adu.edu.tr/fakulte/muhendislik/default.asp)**

**T.C. AYDIN ADNAN MENDERES UNIVERSITY**

**FACULTY OF ENGINEERING**

**DEPARTMENT OF COMPUTER ENGINEERING**

**CSE402 Graduation Thesis 2,Spring 2020**

**Supervisor:Dr.Öğr.Üyesi Mahmut SİNECEN**

**“Disease Diagnosis with Image”**

**Final Report**

**(Bachelor of Science Thesis)**

**22.06.2020**

**By:**

**Seda ÖZCAN(151805016)**

# PLAGIARISM STATEMENT

This report was written by the group members and in our own words, except for quotations from published and unpublished sources which are clearly indicated and acknowledged as such. We are conscious that the incorporation of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism according to the University Regulations. The source of any picture, graph, map or other illustration is also indicated, as is the source, published or unpublished, of any material not resulting from our own experimentation, observation or specimen collecting.

**Project Group Members:**

|  |  |  |  |
| --- | --- | --- | --- |
| Name, Lastname | Student Number | Signature | Date |
| Seda ÖZCAN | 151805016 |  | 22.01.2020 |

**Project Supervisors:**

|  |  |  |  |
| --- | --- | --- | --- |
| Name, Lastname | Department | Signature | Date |
| Mahmut SİNECEN | Computer Engineering |  |  |

# ACKNOWLEDGEMENTS

In my thesis, I would like to thank Mahmut Sinecen for the help, contributions and guidance of our teacher.

# KEYWORDS

#SYMPTOMS

'back\_pain','constipation','abdominal\_pain','diarrhoea','mild\_fever','yellow\_urine',

'yellowing\_of\_eyes','acute\_liver\_failure','fluid\_overload','swelling\_of\_stomach',

'swelled\_lymph\_nodes','malaise','blurred\_and\_distorted\_vision','phlegm','throat\_irritation',

'redness\_of\_eyes','sinus\_pressure','runny\_nose','congestion','chest\_pain','weakness\_in\_limbs',

'fast\_heart\_rate','pain\_during\_bowel\_movements','pain\_in\_anal\_region','bloody\_stool',

'irritation\_in\_anus','neck\_pain','dizziness','cramps','bruising','obesity','swollen\_legs',

'swollen\_blood\_vessels','puffy\_face\_and\_eyes','enlarged\_thyroid','brittle\_nails',

'swollen\_extremeties','excessive\_hunger','extra\_marital\_contacts','drying\_and\_tingling\_lips',

'slurred\_speech','knee\_pain','hip\_joint\_pain','muscle\_weakness','stiff\_neck','swelling\_joints',

'movement\_stiffness','spinning\_movements','loss\_of\_balance','unsteadiness',

'weakness\_of\_one\_body\_side','loss\_of\_smell','bladder\_discomfort','foul\_smell\_of urine',

'continuous\_feel\_of\_urine','passage\_of\_gases','internal\_itching','toxic\_look\_(typhos)',

'depression','irritability','muscle\_pain','altered\_sensorium','red\_spots\_over\_body','belly\_pain',

'abnormal\_menstruation','dischromic \_patches','watering\_from\_eyes','increased\_appetite','polyuria','family\_history','mucoid\_sputum',

'rusty\_sputum','lack\_of\_concentration','visual\_disturbances','receiving\_blood\_transfusion',

'receiving\_unsterile\_injections','coma','stomach\_bleeding','distention\_of\_abdomen',

'history\_of\_alcohol\_consumption','fluid\_overload','blood\_in\_sputum','prominent\_veins\_on\_calf',

'palpitations','painful\_walking','pus\_filled\_pimples','blackheads','scurring','skin\_peeling',

'silver\_like\_dusting','small\_dents\_in\_nails','inflammatory\_nails','blister','red\_sore\_around\_nose',

'yellow\_crust\_ooze']

#DISEASE

'Fungal infection','Allergy','GERD','Chronic cholestasis','Drug Reaction',

'Peptic ulcer diseae','AIDS','Diabetes','Gastroenteritis','Bronchial Asthma','Hypertension',

' Migraine','Cervical spondylosis',

'Paralysis (brain hemorrhage)','Jaundice','Malaria','Chicken pox','Dengue','Typhoid','hepatitis A',

'Hepatitis B','Hepatitis C','Hepatitis D','Hepatitis E','Alcoholic hepatitis','Tuberculosis',

'Common Cold','Pneumonia','Dimorphic hemmorhoids(piles)',

'Heartattack','Varicoseveins','Hypothyroidism','Hyperthyroidism','Hypoglycemia','Osteoarthristis',

'Arthritis','(vertigo) Paroymsal Positional Vertigo','Acne','Urinary tract infection','Psoriasis',

'Impetigo'

**Respirate rate=**The respiration rate is the number of breaths a person takes per minute. The rate is usually measured when a person is at rest and simply involves counting the number of breaths for one minute by counting how many times the chest rises.

**Blood pressure=**Blood pressure inside the arteries is called blood pressure. When the kapl muscle contracts, blood is pumped into the arteries. The pressure measured at this time is the systolic pressure. When the heart muscle relaxes, blood is stopped in the vein; meanwhile, the measured pressure is diastolic pressure

**Cholestrol=**Cholesterol is a waxy, fat-like substance that your body needs. But when you have too much in your blood, it can build up on the walls of your arteries and form blockages. This can lead to heart disease, heart attack, and stroke.

**Pulsation=** the [action](https://dictionary.cambridge.org/dictionary/english/action) of [moving](https://dictionary.cambridge.org/dictionary/english/moving) with a [strong](https://dictionary.cambridge.org/dictionary/english/strong), [regular](https://dictionary.cambridge.org/dictionary/english/regular) [rhythm](https://dictionary.cambridge.org/dictionary/english/rhythm):

*The star's pulsations*[*cause*](https://dictionary.cambridge.org/dictionary/english/cause)*it to*[*brighten*](https://dictionary.cambridge.org/dictionary/english/brighten)*and*[*dim*](https://dictionary.cambridge.org/dictionary/english/dim)*every six*[*minutes*](https://dictionary.cambridge.org/dictionary/english/minutes)*.*

*Every*[*blink*](https://dictionary.cambridge.org/dictionary/english/blink)*of an*[*eye*](https://dictionary.cambridge.org/dictionary/english/eye)*or pulsation of a paramecium is*[*powered*](https://dictionary.cambridge.org/dictionary/english/power)*by the*[*protein*](https://dictionary.cambridge.org/dictionary/english/protein)[*ATP*](https://dictionary.cambridge.org/dictionary/english/atp)*.*

# ABSTRACT

In this thesis project, it is aimed to facilitate the pre-diagnosis of patients coming to the emergency department and to decrease the mortality rate caused by misdiagnosis in the emergency department.

For preliminary diagnosis, triage method is applied first and patient urgency is graded. The patient whose urgency was detected was grouped as green, yellow and red areas.

Grouping and preliminary diagnosis; It will be performed by evaluating the image to be taken as soon as the patient arrives at the emergency room and vital values ​​to be entered manually.

I tried to create a fuzzy logic model by using python programming language and matlab application in order to get colored region output by evaluating the ratios of vital values. I used random data in the fuzzy logic model I created.

I designed an interface to collect real data.

Vital values ​​entered in the interface, the symptoms seen and patient images to be taken will make estimates by combining the results with machine learning algorithms by separately estimating at the first stage.

# ÖZET

Bu tez projesinde, acil servise gelen hastaların ön tanısının kolaylaştırılması ve acil serviste yanlış tanı konmasının neden olduğu mortalite oranının düşürülmesi amaçlanmıştır.

Ön tanı için önce triyaj yöntemi uygulanır ve hastanın aciliyeti derecelendirilir. Aciliyeti saptanan hasta yeşil, sarı ve kırmızı alanlar olarak gruplandırılmıştır.

Gruplama ve ön tanı; Hasta acil servise gelir gelmez alınacak görüntüyü ve manuel olarak girilecek hayati değerleri değerlendirerek gerçekleştirilecektir.

Vital değerlerin oranları değerlendirilerek renkli bölge çıktısı alabilmek için python programlama dili ve matlab uygulamasını kullanarak fuzzy logic model oluşturmaya çalıştım. Oluşturduğum fuzzy logic modelinde random veriler kullandım.

Gerçek verileri toplamak için bir arayüz tasarımı yaptım.

Arayüzde girilen vital değerler, görülen semptomlar ve alınacak hasta görüntüleri ilk etapta ayrı ayrı tahminler de bulunarak machine learning algoritmalarıyla birlikte sonuçları birleştirerek hastalık tahminleri yapacak

# TABLE OF CONTENTS

[PLAGIARISM STATEMENT ii](#_Toc512447154)

[ACKNOWLEDGEMENTS iii](#_Toc512447155)

[KEYWORDS iv](#_Toc512447156)

[ABSTRACT v](#_Toc512447157)

[ÖZET vi](#_Toc512447158)

[TABLE OF CONTENTS vii](#_Toc512447159)

[LIST OF FIGURES viii](#_Toc512447160)

[LIST OF TABLES ix](#_Toc512447161)

[LIST OF ACRONYMS/ABBREVIATIONS x](#_Toc512447162)

[1. INTRODUCTION 1](#_Toc512447163)

[1.1. Description of the Problem 1](#_Toc512447164)

[1.2. Project Goal 1](#_Toc512447165)

[1.3. Project Output 1](#_Toc512447166)

[1.4. Project Activities and Schedule 1](#_Toc512447167)

[2. DESIGN 1](#_Toc512447168)

[2.1. High Level Design 1](#_Toc512447169)

[2.2. Detailed Design 1](#_Toc512447170)

[2.3. Realistic Restrictions and Conditions in the Design 2](#_Toc512447171)

[3. IMPLEMENTATION, TESTS and TEST DISCUSSIONS 2](#_Toc512447172)

[3.1. Implementation of the Product 2](#_Toc512447173)

[3.2. Tests and Results of Tests 2](#_Toc512447174)

[4. CONCLUSIONS 2](#_Toc512447175)

[4.1. Summary 2](#_Toc512447176)

[4.2. Cost Analysis 2](#_Toc512447177)

[4.3. Benefits of the Project 2](#_Toc512447178)

[4.4. Future Work 3](#_Toc512447179)

[References 3](#_Toc512447180)

[APPENDICES 4](#_Toc512447181)

[APPENDIX A: REQUIREMENTS SPECIFICATION DOCUMENT 5](#_Toc512447182)

[APPENDIX B: DESIGN SPECIFICATION DOCUMENT 6](#_Toc512447183)

[APPENDIX C: PRODUCT MANUAL 7](#_Toc512447184)

[APPENDIX D: SOURCE CODE/EXECUTABLES/SCRIPTS IN CD/DVD 8](#_Toc512447185)

# 1. INTRODUCTION

# 1.1. Description of the Problem

Today, hundreds of people lose their lives due to being late in the intervention of the patients coming to the emergency department or due to the wrong pre-diagnosis.

I believe we can solve this problem with this project

# 1.2. Project Goal

In this thesis project, it is aimed to facilitate the preliminary diagnosis of patients coming to the emergency room and to decrease the mortality rate caused by misdiagnosis in the emergency room.

# 1.3. Project Output

I am going to create a software product for this purpose of project.

# Used programs

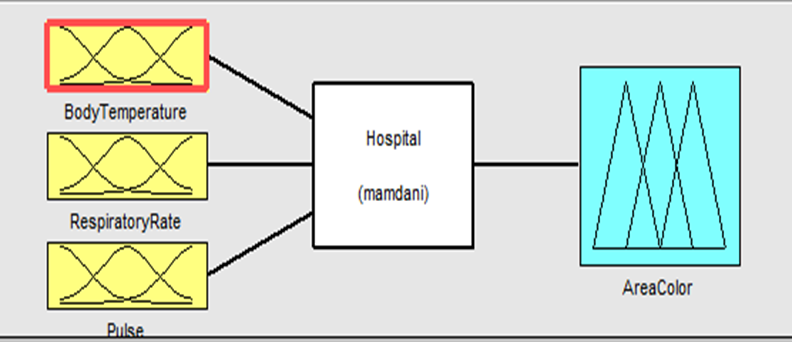
* Python
* Matlab(fuzzy model)
* Jupyter Notebook(fuzzy)
* Spyder
* QTDesigner
* Tkinter
* Visual Studio2010
* SQLite Database

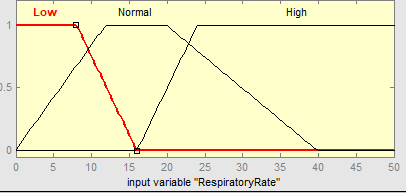
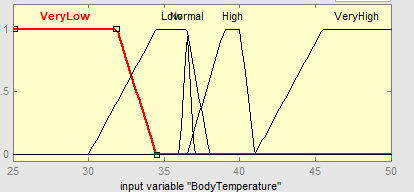
# 

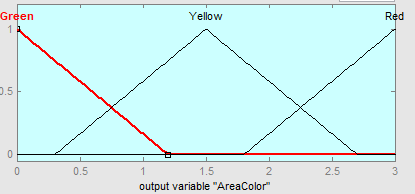
# 2. DESIGN

In my project, First of all, I created a fuzzy logic model with random data and tried to make colored area determination for triage application, but this only rates patient urgency.

FIGURE 1:Design of fuzzy model which ı created



****

****FIGURE2:respiratory rate on fuzzy model FIGURE3:body temperature on fuzzy m

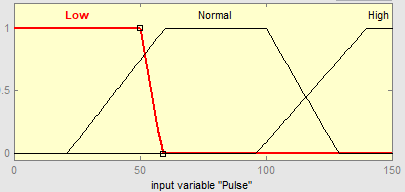
****

FIGURE4:Pulse on fuzzy model FIGURE5:Output Area colour on fuzzy m.

Then, I create an interface with PyQt5 and tkinter.

I use QT Designer and spyder

In my interface user will create account and login. When the user login, it will enter main interface that will enter vital values and upload images.

.

FIGURE6:Login interface



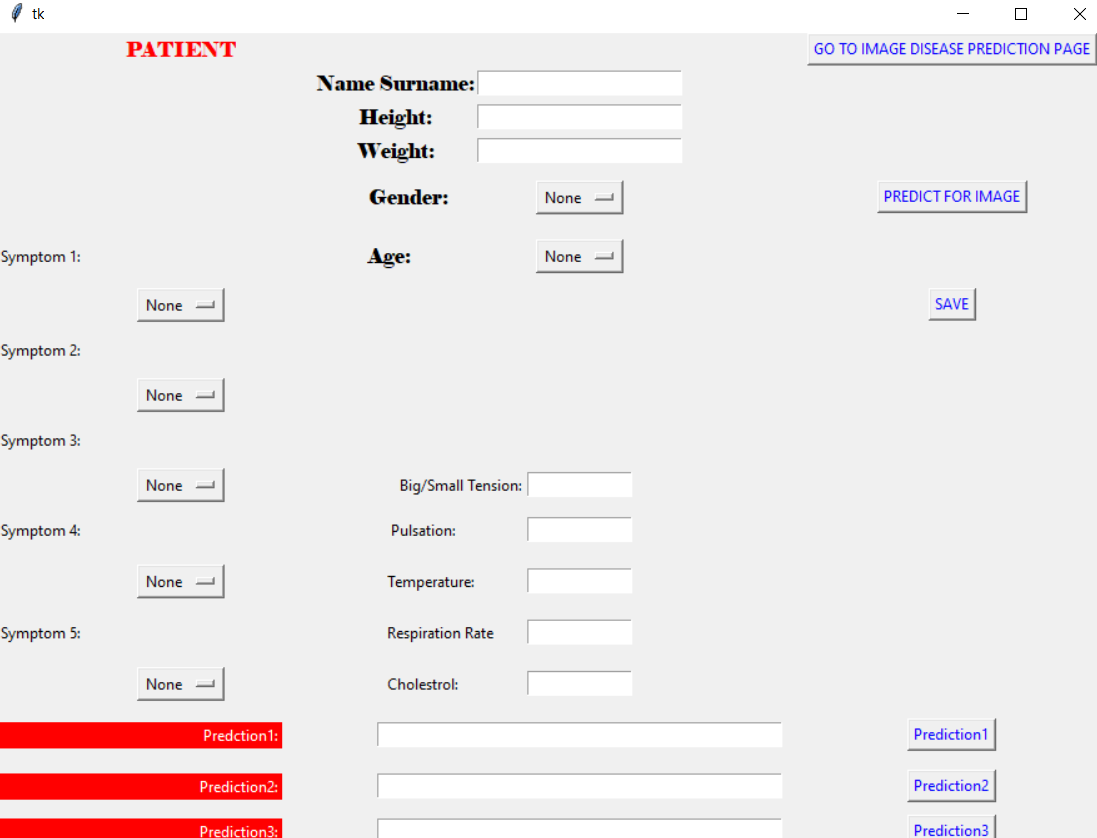
FIGURE7:Create account interface

My project consists of two parts, image processing and machine algorithms

I use 3 machine learning algorithm (Decision tree, Random Forest, Naïve Bayes)

These algorithms are working based on symptomps.

FIGURE8:Main interface



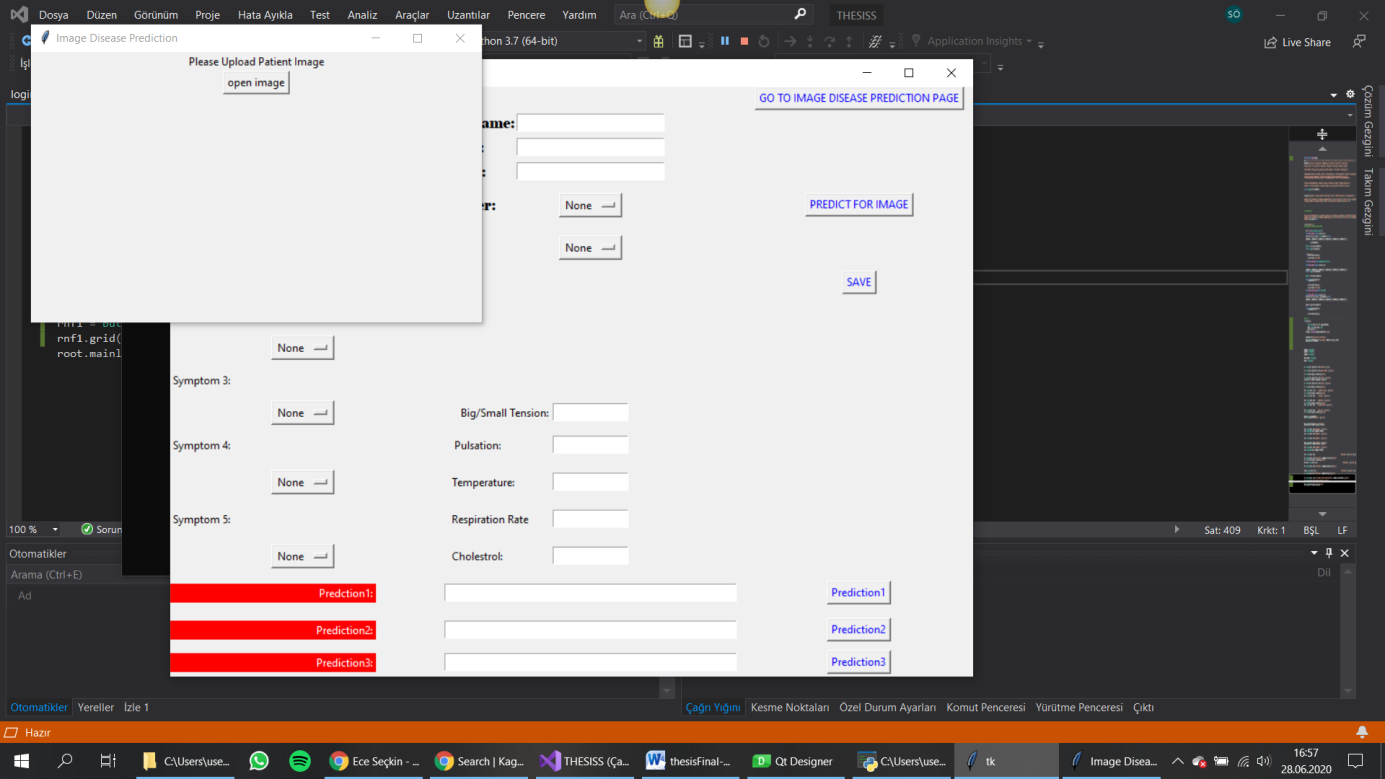


FIGURE9:Open İmage

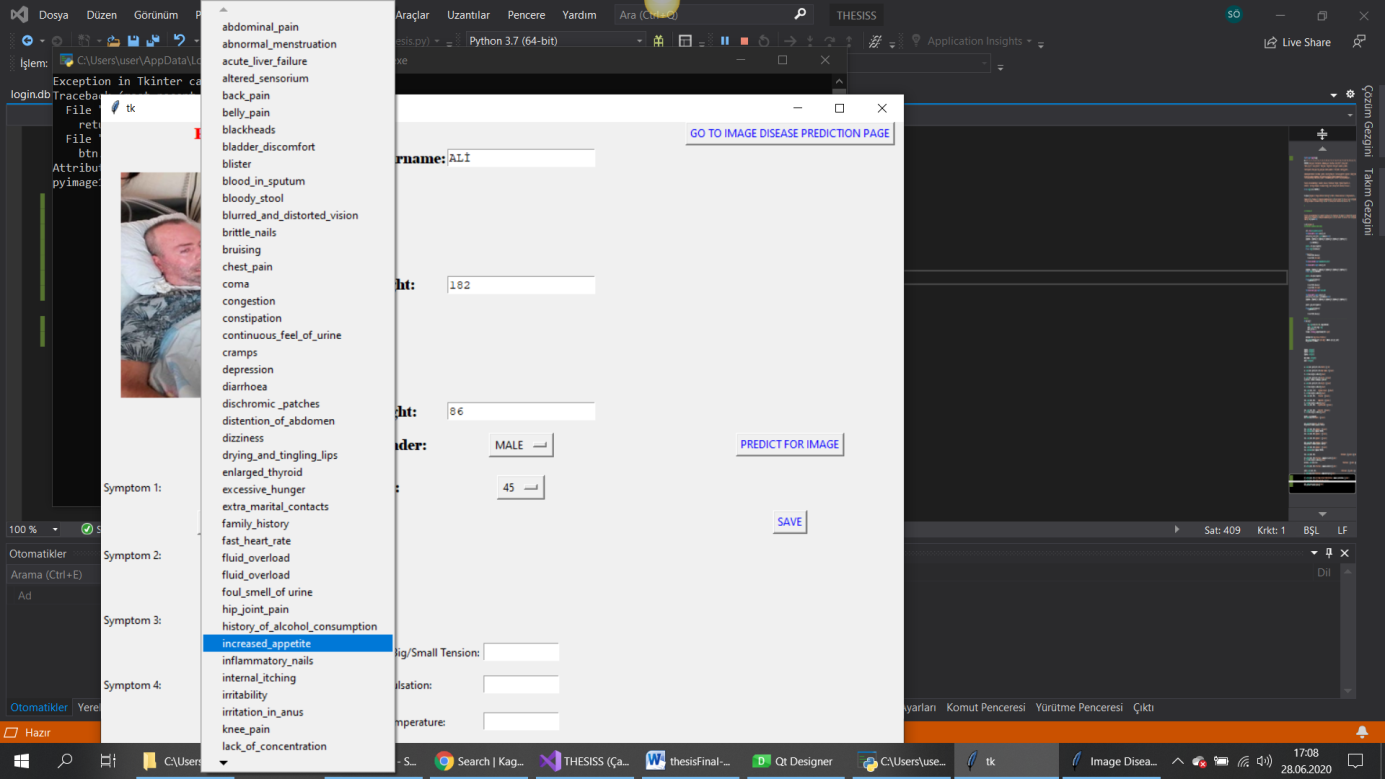
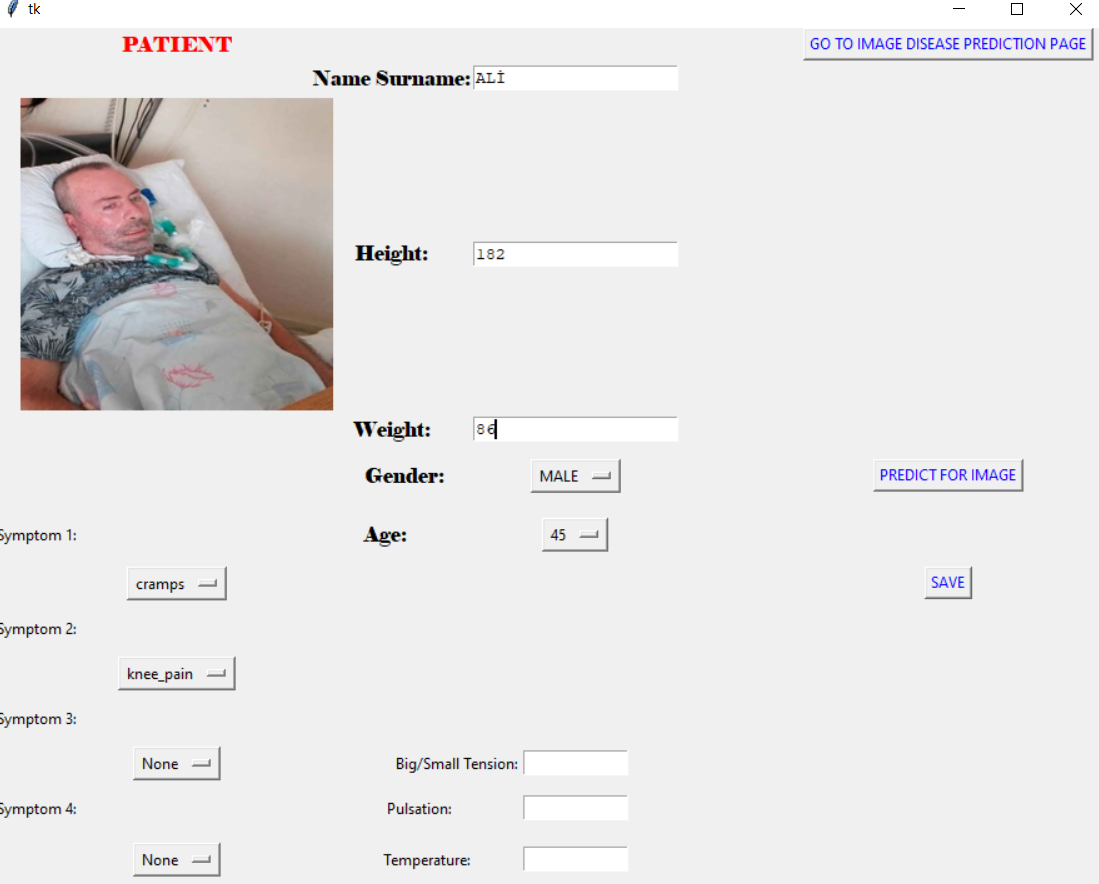
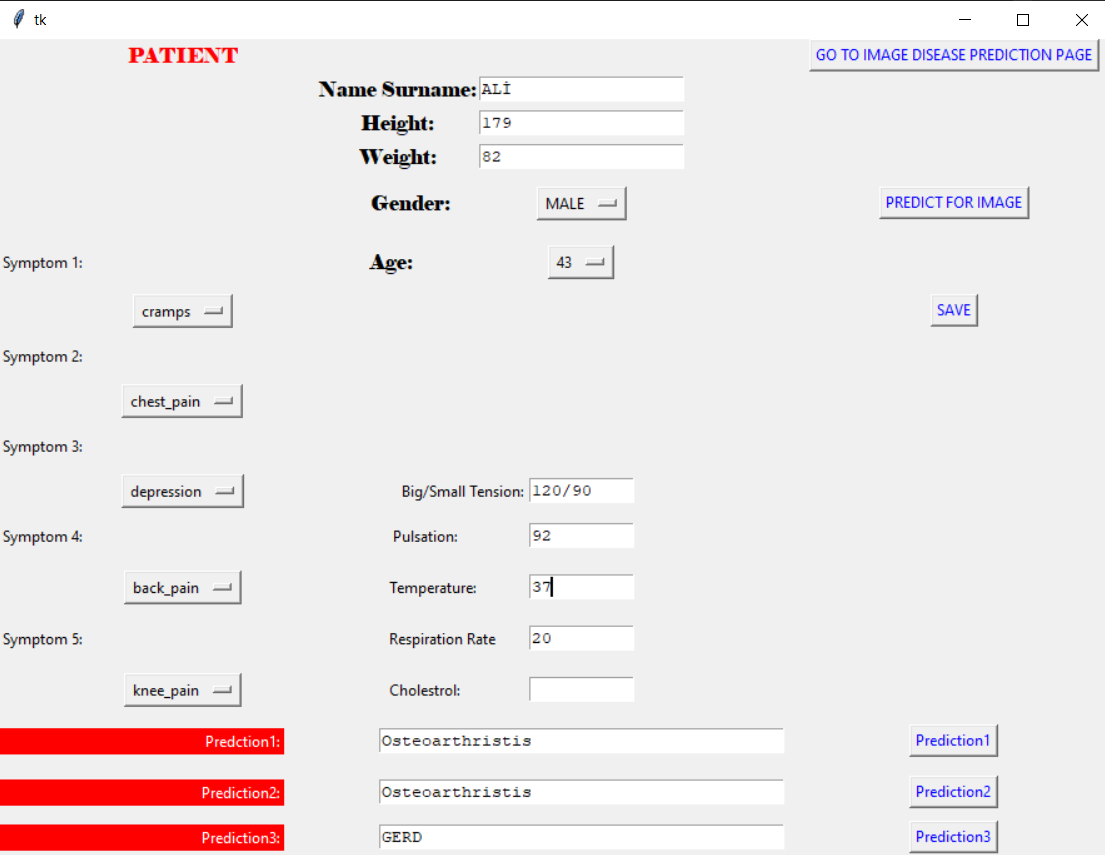


FIGURE10: Symptomps





Prediction 1 analyze the symptomps for decision tree

Prediction 2 analyze the symptomps for random forest

Prediction 3 analyze the symptomps for naïve bayes algorithm.

# 4. CONCLUSIONS

# 4.1. Summary

Although the importance of human health is very important today, the risk of death of patients requiring emergency treatment in a sudden emergency situation is greater than other patients. In this field, emergency departments of all hospitals and health institutions serve. The emergency service is the unit that provides this service to patients in need of emergency health assistance in an area where hospitals and other health institutions can easily access and enter the ambulances. Triage Application is used to classify emergency situations and accelerate the interventions in emergency departments. Color coding is applied for effective service delivery in the emergency unit. Triage application is performed by physician or nurse, health officer (community health) and similar health personnel during emergency application. For triage application, red, yellow and green colors are used in order of priority in terms of examination, examination, treatment, medical and surgical interventions.

Preliminary diagnosis is the cornerstone for the treatment of the patient. If this happens easily and healthily, the patient's recovery time decreases and the efficiency from the treatment increases.

In this project, which we are trying to develop to speed up the pre-diagnosis and treatment process, we will introduce a software product that provides efficient results using fuzzy model, image processing and machine learning algorithms.

# 4.3. Benefits of the Project

Thanks to this project

* To determine the risk group of patients coming to the emergency department
* To accelerate the first intervention and referral of these patients
* To facilitate the emergency department team to perform their duties
* To reduce the number of patients lost due to late and incorrect intervention in the emergency department
* to diagnose patients easily

will become possible.

# 4.4. Future Work

In the continuation of the project: real images and real vital data will be collected with the interface. With the image processing, the data collected from the image and the vital data taken manually from the interface will be blended preliminary diagnosis study will be done.

* I will collect real image datas
* I will complete image process part of project
* I will test and develop to reach most qualified and useful version of project.
* I will save delete mods for results in this platform.

**LIMITATIONS and REVİEWS**

I couldn’t create image processin and image data collect.

And I couldn’t work real datas.

I create a fuzzy model but I did not get healthy results

In the other hand, Machine Learning algorithms based on symptoms worked more healthy

I think that if I can develop and make this project the most efficient, it will be a great convenience for the health sector in our country and in the world and will support human health. and then I want to continue developing this project.

# References

* <http://vucut-sicakligi.nedir.org/>
* <https://www.acibadem.com.tr/Hayat/Bilgi/yasam-bulgulari-ile-ilgili-onemli-gostergeler-nelerdir>
* <https://www.medicalpark.com.tr/nabiz-nedir-ideal-nabiz-kac-olmali/hg-1941>
* <https://yedikulegoguseah.saglik.gov.tr/TR,63011/acil-saglik-hizmetleri.html>
* <https://www.acibadem.com.tr/Hayat/Bilgi/yasam-bulgulari-ile-ilgili-onemli-gostergeler-nelerdir>
* <https://www.haberturk.com/saglik/haber/937086-acil-servislere-renkli-alan-uygulamasi>
* <https://github.com/Vnitjain/Washing-Machine-Fuzzy-Controller/blob/master/wmfuzzy.py>
* <https://github.com/atasoglu98/Fuzzy-Brake-Example>
* <https://medium.com/deep-learning-turkiye/python-ile-bulan%C4%B1k-mant%C4%B1k-modellemesi-74459dc27308>?
* <http://acikerisim.pau.edu.tr:8080/xmlui/bitstream/handle/11499/1844/Mahmut%20Sinecen.doc?sequence=1&isAllowed=y>
* <https://www.toraks.org.tr/halk/News.aspx?detail=4013>
* <https://www.youtube.com/watch?v=Bym5KIEIvHY&list=PLK6Whnd55IH5rbIrl61H2l-560LddyHO5&index=39>
* <https://www.youtube.com/watch?v=oJJIV6bBbNo>
* <http://fbe.cu.edu.tr/tr/makaleler/2000/UNSUPERVISED.pdf>
* FUZZY LOGIC WITH ENGINEERING APPLICATIONS (Timothy J. ROSS)
* (Third edition/2010)
* <https://www.smarttekas.com.tr/felcli-hasta-ile-iletisim/>
* <https://prohealthsys.com/central/assessment/physical-assessment/vital-signs/vital_signs_table/>
* <https://www.geeksforgeeks.org/loading-images-in-tkinter-using-pil/>
* <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
* <https://blog.easysol.net/machine-learning-algorithms-4/>
* <https://breathedc.org/cholesterol-and-lung-health/>
* <https://www.medicalpark.com.tr/tansiyon-nedir-cesitleri-ve-dikkat-edilmesi-gerekenler-nelerdir/hg-1761#:~:text=Atardamarlar%C4%B1n%20i%C3%A7indeki%20kan%20bas%C4%B1nc%C4%B1na%20tansiyon,esnada%20%C3%B6l%C3%A7%C3%BClen%20bas%C4%B1n%C3%A7%20diastolik%20bas%C4%B1n%C3%A7t%C4%B1r>.
* <https://dictionary.cambridge.org/dictionary/english/pulsation>

# APPENDICES

#RESEARCHES#

**How do physicians diagnose?**

Unfortunately, this is not as easy as I think, and sometimes it can take days, even months, to diagnose. There are 100 thousand diseases. The physician tries to reduce this number to a reasonable number with the questions asked to the patient, necessary examinations and examinations.

If somebody is obviously among the possibilities, he / she would prefer to be treated for it. If more than one possibility still appears with equal probability, further investigations may be required.

**STRUCTURAL ARRANGEMENT OF EMERGENCY SERVICE**

Emergency health services unit; patient-counseling, patient registration, triage, patient examination room, resuscitation room, observation rooms, isolation room, emergency surgery room, waiting area, physician, nurse and other auxiliary personnel rooms.

EMERGENCY DEPARTMENT PATIENT ACCEPTANCE PROCESSES

PATIENT RECOMMENDATION AND GUIDANCE

• The patients and their relatives who are admitted to the emergency department are welcomed and accompanied by the patient if they need to be directed. Patient reception and counseling staff receive interpersonal communication, ethics, patient rights and responsibilities, and patient satisfaction trainings at least once a year.

EMERGENCY POLYCLINIC PATIENT ADMISSION

• Patients coming to the Emergency Health Services Unit by ambulance, 112 and stretcher are taken directly to the yellow room. Those who are in critical condition and who are in danger of life are taken to the red room. Other applications are forwarded to the triage area prior to registration.

• Necessary emergency procedures of the patients evaluated by the doctor in medical terms are initiated. Barcodes are taken out by the emergency secretary after they are registered to HIS. The procedures of patients without relatives are carried out by emergency personnel.

• Prescriptions of patients who are examined in the green room are given. Patients whose general condition requires good and advanced examination are referred to the polyclinic.

• The patients who have been directed to the yellow room and whose medical examination and intervention have been initiated are treated and discharged.

• Emergency medical examination and interventions directed to the yellow room are initiated and the patients who are given a decision of hospitalization as a result of medical evaluation are given white ID (red color for patients with allergies).

• Patients are admitted directly to the ward or intensive care unit according to their medical condition. If there is no empty bed in the hospital, the patient is referred to other health institutions with the coordination of 112.

• General Forensic Examination Report is filled and stamped and signed by the patient's doctor for the trauma and forensic patients who apply to the emergency department. For patients covered by forensic cases, the police are informed. The patient's arrival time, identity information and reason of arrival are recorded in the forensic case section on the computer. The forensic case file is created and archived. A copy of the report is delivered to the police.

TRIAGE APPLICATION

• Color coding is applied for effective service delivery in the emergency unit.

• Triage application is performed by the physician or nurse, health officer (community health) and similar health personnel during the emergency application.

• For triage application, red, yellow and green colors are used in order of priority in terms of examination, examination, treatment, medical and surgical interventions. Application of color coding;

• Green area: Outpatients are not critically injured, can walk and take care of themselves. NORMAL PATIENTS Intervention is required within 24 hours.

• Yellow area: Patients in this group may be kept a little longer than those in the red group. The delay in caring for a patient who needs medical care but is caring for other patients will not cause that patient to be ex-affected. Yellow patients are not outpatients and need a stretcher for transplant. SECOND EMERGENCY PATIENTS should be treated within 60 minutes.

• Red area: Emergency group. Critically seriously injured patients whose problems or injuries need to be treated promptly. In this group, priority is given to those who need to receive or transfer health services. PRIMARY EMERGENCY PATIENTS and the intervention is done immediately. Critical patients receive immediate medical attention regardless of whether or not registration is performed. Necessary arrangements regarding automation system have been provided in order to make the registration of critical patients primarily.

1. Triage procedures of all patients admitted to the emergency department are completed as soon as possible. Following this, their records are opened, they are put in line for diagnosis and treatment procedures taking into consideration their medical conditions and taken to the area that is in accordance with the triage code.

APPLICATION FOR PATIENT REGISTRATION

• All patients who apply to the Emergency Health Services Unit (patients who apply to the red and yellow areas) are accepted regardless of health insurance and solvency. For the cases to be brought to the unit by 112 emergency ambulances, the emergency department is prepared, first evaluation and necessary medical intervention is performed and the patient is stabilized. In this context, the contents of Sunum Circular of Emergency Health Services Presentation ve and Provincial Emergency Health Services “ASKOM” decisions are fully implemented.

• The records of the patients who apply to the Emergency Health Services Unit make the patient admission record together with the identity check by the emergency secretary and the social security inquiry. Every patient applying to the Emergency Health Services Unit is recorded. Emergency (red) patients are intervened without questioning their identity. Operations are then completed.

• A “Patient Information Form” is filled in to determine the identity information and address of the patients who apply to the emergency services but who cannot submit a valid identity document, health card or patient referral documents for any reason. No deed or undertaking is taken from these patients.

• Reception of foreign patients; The passport photocopy and ID photocopy of the patient is made by the cashier unit and the patient is registered after the treatment is made. No fee is paid to patients who are documented as forensic foreigners and refugees.

• For Syrian patients; According to the Prime Ministry circular, Syrian patients are registered with their passport and ID photocopy. There is no charge from the patient without any treatment and application.

PATIENT EXAMINATION, DIAGNOSIS AND INTERVENTION PROCEDURES

• Patients who are admitted to the red room are treated immediately. Stabilizers are admitted to the service or intensive care unit, depending on the situation. When necessary, it is referred to other health institutions with the coordination of 112. Ex patients are notified to the officer in charge and procedures are initiated.

• The patient, who is examined in the yellow room, is taken for observation. Diagnosis and treatment examinations are carried out by informing physicians and nurses.

• For laboratory tests, blood sample is taken by the nurse and blood gas is taken by the physician. The examinations are processed by the secretary in automation and delivered to the laboratory by the personnel.

• The personnel takes the patient to the emergency radiology department for the required radiological examinations in the emergency unit. The results are printed in HIS system or in 15 minutes at the latest.

• Patients whose medical condition is critical during imaging and similar procedures are accompanied by a health personnel as required by their condition. If the patient needs it, they are sent with a stretcher or wheelchair.

• The findings of the patients evaluated in the yellow room are recorded in the Ambulatory Patient Evaluation and Follow-up Form. For each patient undergoing an examination, the patient is authenticated before any procedure.

• An examination may be requested from patients who are examined in the green room. For the examinations requested by the physician, they are directed to the relevant data entry element and their prescriptions are given after the results are evaluated. Patients whose general condition requires good and advanced examination are referred to the polyclinic.

What is Triage?

It refers to the process of prioritization by medical or trained medical personnel considering the complaints, severity of symptoms and urgency of medical conditions of patients admitted to emergency services. Triage is done at the time of application. For triage application, red, yellow and green colors are used in order of priority in terms of examination, examination, treatment, medical and surgical interventions.

Green area: An outpatient is a simple state of health that is stable in general and can be treated as an outpatient.

They are outpatients and are not critically injured, can walk and take care of themselves. NORMAL PATIENTS Intervention is required within 24 hours.

Yellow Area: Life-threatening conditions, risk of limb loss and significant morbidity rates, with moderate to prolonged symptoms and potentially serious.

Patients in this group may be kept a little longer than those in the red group. The delay in caring for a patient who needs medical care but is caring for other patients will not cause that patient to be ex-affected. Yellow patients are not outpatients and need a stretcher for transplant. SECOND EMERGENCY PATIENTS should be treated within 60 minutes.

Red Area: Life-threatening and rapid aggressive approach and situations requiring immediate simultaneous assessment and treatment. In these cases, the patient is taken to the red area without waiting. They are also life-threatening situations that need to be evaluated and treated within 10 minutes.

Emergency group. Critically seriously injured patients whose problems or injuries need to be treated promptly. In this group, priority is given to those who need to receive or transfer health services. PRIMARY EMERGENCY PATIENTS and the intervention is done immediately. Critical patients receive immediate medical attention regardless of whether or not registration is performed. Necessary arrangements regarding automation system have been provided in order to make the registration of critical patients primarily.

1. Triage procedures of all patients admitted to the emergency department are completed as soon as possible. Following this, their records are opened, they are put in line for diagnosis and treatment procedures taking into consideration their medical conditions and taken to the area that is in accordance with the triage code.

What are the important indicators of life signs

Before evaluating the patient / injured person during first aid, the meaning of life signs should be known. Because; these findings are important for interventions. Life signs of the patient / injured;

Awareness,

Respiratory,

Circulation,

Body temperature,

Refers to blood pressure.

Assessment of consciousness:

A normal person responds to all warnings directed to him / her. The level of consciousness indicates the severity of the injury.

If the person is conscious, he / she responds to all warnings.

In the first degree loss of consciousness; verbal and noisy stimuli,

In the second degree of unconsciousness; painful stimuli,

Third degree loss of consciousness; it is unresponsive to all stimuli and does not respond.

 Respiratory assessment:

When assessing the patient's / injured's breathing, the frequency of breathing, the equality of breathing intervals, the depth of breathing are examined. In a healthy adult person, the respiratory rate is 12–20 per minute, 16–22 in children, and 18’24 in infants.

Blood pressure assessment:

Blood pressure is not controlled when evaluating the patient / injured. However, knowing the meaning of blood pressure

It is the pressure on the vessel wall during contraction and relaxation of the heart. The blood pumping power of the heart

Shows. The normal value is 100 / 50-140/100 mm Hg.

 Heart rate assessment:

The pressure of the heart beats on the arterial wall can be felt with fingertips on the vessel wall.

It is called pulse.

In an adult, the normal pulse rate is 60–100 per minute,

100–120 in children,

100-140 in babies.

Pulse areas in the body; carotid artery, forearm vein, leg vein, arm vein. When evaluating the circulation of the patient / injured, pulse is taken from the jugular vein in children and adults and from the arm artery in infants.

Assessment of body temperature:

In first aid, body temperature should be measured under the armpit. Normal body temperature is 36.5 degrees. Above normal value is indicated as high fever, below is indicated as low fever. Indicates a danger above 41–42 degrees and 34.5 degrees. 31.0 degrees and below is fatal.

FUZZY LOGIC

Fuzzy logic is a very valuable form of logic derived from fuzzy set theory related to logical analysis.

This is not certain, it is an approximate value. In fuzzy logic, the degree of accuracy of an expression

0 to 1 and is not limited to two accuracy values, such as {0,1} or {false, true}

Classical predicate logic (Huang, Jo, Lee, Kang and Bevilacqua, 2009). Creating a basis for

approximate reasoning, that is, a type of reasoning that is not exactly what it is,

logic can provide a more realistic framework for human reasoning than traditional two assessors

logic (Zadeh, 1975). The term "fuzzy logic" emerged as a result of this development.

Lotfi Zadeh's theory of fuzzy sets. In 1965, Lotfi Zadeh proposed fuzzy set theory (Zadeh,

1965) and later created fuzzy logic based on fuzzy sets.

MACHINE LEARNING ALGORITHMS

### Introduction

Classification is a two-step process, learning step and prediction step, in machine learning. In the learning step, the model is developed based on given training data. In the prediction step, the model is used to predict the response for given data. Decision Tree is one of the easiest and popular classification algorithms to understand and interpret.

### Decision Tree Algorithm

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving **regression and classification problems** too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by **learning simple decision rules** inferred from prior data(training data).

In Decision Trees, for predicting a class label for a record we start from the **root** of the tree. We compare the values of the root attribute with the record’s attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

### Types of Decision Trees

Types of decision trees are based on the type of target variable we have. It can be of two types:

1. **Categorical Variable Decision Tree:**Decision Tree which has a categorical target variable then it called a **Categorical variable decision tree.**
2. **Continuous Variable Decision Tree:**Decision Tree has a continuous target variable then it is called **Continuous Variable Decision Tree.**

**Example:-** Let’s say we have a problem to predict whether a customer will pay his renewal premium with an insurance company (yes/ no). Here we know that the income of customers is a significant variable but the insurance company does not have income details for all customers. Now, as we know this is an important variable, then we can build a decision tree to predict customer income based on occupation, product, and various other variables. In this case, we are predicting values for the continuous variables.

### Important Terminology related to Decision Trees

1. **Root Node:**It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
2. **Splitting:**It is a process of dividing a node into two or more sub-nodes.
3. **Decision Node:**When a sub-node splits into further sub-nodes, then it is called the decision node.
4. **Leaf / Terminal Node:**Nodes do not split is called Leaf or Terminal node.
5. **Pruning:**When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
6. **Branch / Sub-Tree:**A subsection of the entire tree is called branch or sub-tree.
7. **Parent and Child Node:**A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.



Decision trees classify the examples by sorting them down the tree from the root to some leaf/terminal node, with the leaf/terminal node providing the classification of the example.

Each node in the tree acts as a test case for some attribute, and each edge descending from the node corresponds to the possible answers to the test case. This process is recursive in nature and is repeated for every subtree rooted at the new node.

### Assumptions while creating Decision Tree

Below are some of the assumptions we make while using Decision tree:

* In the beginning, the whole training set is considered as the **root.**
* Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
* Records are **distributed recursively** on the basis of attribute values.
* Order to placing attributes as root or internal node of the tree is done by using some statistical approach.

Decision Trees follow **Sum of Product (SOP) r**epresentation. The Sum of product (SOP) is also known as **Disjunctive Normal Form**. For a class, every branch from the root of the tree to a leaf node having the same class is conjunction (product) of values, different branches ending in that class form a disjunction (sum).

The primary challenge in the decision tree implementation is to identify which attributes do we need to consider as the root node and each level. Handling this is to know as the attributes selection. We have different attributes selection measures to identify the attribute which can be considered as the root note at each level.

### How do Decision Trees work?

The decision of making strategic splits heavily affects a tree’s accuracy. The decision criteria are different for classification and regression trees.

Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that the purity of the node increases with respect to the target variable. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.

The algorithm selection is also based on the type of target variables. Let us look at some algorithms used in Decision Trees:

**ID3** → (extension of D3)  
**C4.5** → (successor of ID3)  
**CART** → (Classification And Regression Tree)  
**CHAID** → (Chi-square automatic interaction detection Performs multi-level splits when computing classification trees)  
**MARS** → (multivariate adaptive regression splines)

The ID3 algorithm builds decision trees using a top-down [greedy search](https://www.hackerearth.com/practice/algorithms/greedy/basics-of-greedy-algorithms/tutorial/)approach through the space of possible branches with no backtracking. A greedy algorithm, as the name suggests, always makes the choice that seems to be the best at that moment.

**Steps in ID3 algorithm:**

1. It begins with the original set S as the root node.
2. On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates **Entropy(H)** and **Information gain(IG)**of this attribute.
3. It then selects the attribute which has the smallest Entropy or Largest Information gain.
4. The set S is then split by the selected attribute to produce a subset of the data.
5. The algorithm continues to recur on each subset, considering only attributes never selected before.

### Attribute Selection Measures

If the dataset consists of **N** attributes then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step. By just randomly selecting any node to be the root can’t solve the issue. If we follow a random approach, it may give us bad results with low accuracy.

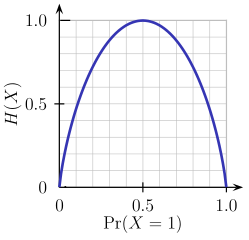
For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some criteria like :

**Entropy**,  
**Information gain,**  
**Gini index,**  
**Gain Ratio,**  
**Reduction in Variance**  
**Chi-Square**

These criteria will calculate values for every attribute. The values are sorted, and attributes are placed in the tree by following the order i.e, the attribute with a high value(in case of information gain) is placed at the root.  
While using Information Gain as a criterion, we assume attributes to be categorical, and for the Gini index, attributes are assumed to be continuous.

### ****Entropy****

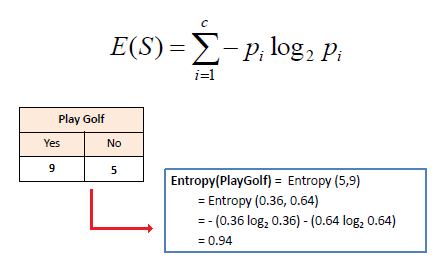
Entropy is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information. Flipping a coin is an example of an action that provides information that is random.



From the above graph, it is quite evident that the entropy H(X) is zero when the probability is either 0 or 1. The Entropy is maximum when the probability is 0.5 because it projects perfect randomness in the data and there is no chance if perfectly determining the outcome.

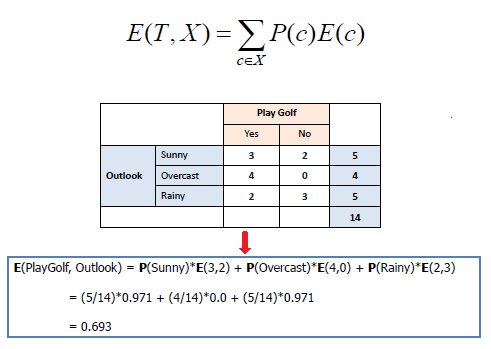
**ID3 follows the rule — A branch with an entropy of zero is a leaf node and A brach with entropy more than zero needs further splitting.**

Mathematically Entropy for 1 attribute is represented as:



Where **S → Current state, and Pi → Probability of an event i of state S or Percentage of class i in a node of state S.**

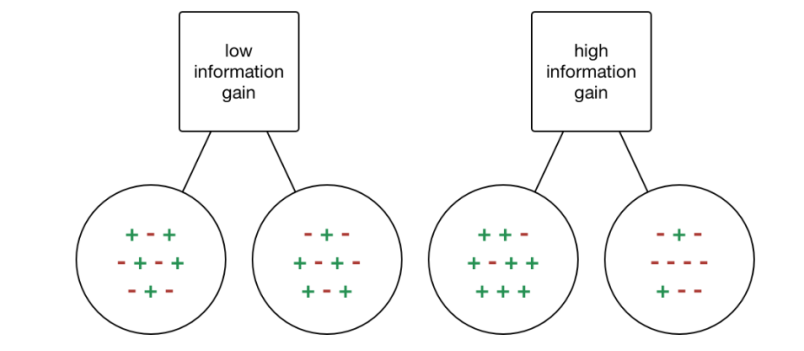
Mathematically Entropy for multiple attributes is represented as:



where**T→ Current state and X → Selected attribute**

### ****Information Gain****

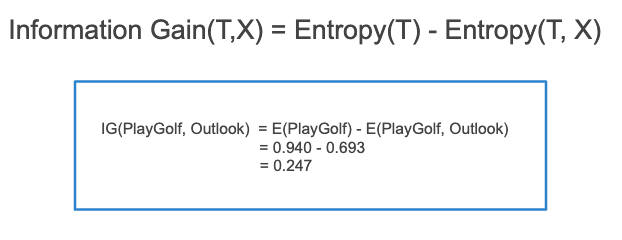
Information gain or **IG**is a statistical property that measures how well a given attribute separates the training examples according to their target classification. Constructing a decision tree is all about finding an attribute that returns the highest information gain and the smallest entropy.



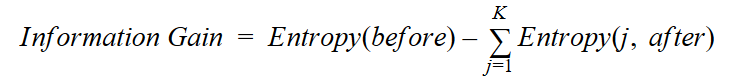
[Information Gain](https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a?gi=a8ffb5170258)

Information gain is a decrease in entropy. It computes the difference between entropy before split and average entropy after split of the dataset based on given attribute values. ID3 (Iterative Dichotomiser) decision tree algorithm uses information gain.

Mathematically, IG is represented as:



In a much simpler way, we can conclude that:

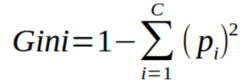


[Information Gain](https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147)

Where “before” is the dataset before the split, K is the number of subsets generated by the split, and (j, after) is subset j after the split.

### Gini Index

You can understand the Gini index as a cost function used to evaluate splits in the dataset. It is calculated by subtracting the sum of the squared probabilities of each class from one. It favors larger parti1tions and easy to implement whereas information gain favors smaller partitions with distinct values.



Gini Index

Gini Index works with the categorical target variable “Success” or “Failure”. It performs only Binary splits.

Higher the value of Gini index higher the homogeneity.

**Steps to Calculate Gini index for a split**

1. Calculate Gini for sub-nodes, using the above formula for success(p) and failure(q) (p²+q²).
2. Calculate the Gini index for split using the weighted Gini score of each node of that split.

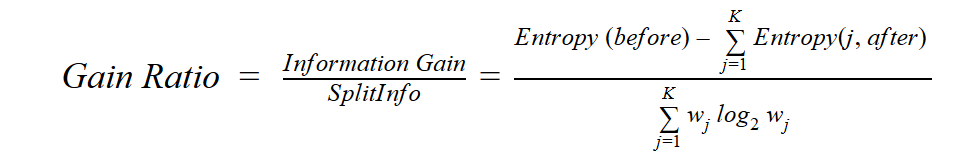
CART (Classification and Regression Tree) uses the Gini index method to create split points.

### Gain ratio

Information gain is biased towards choosing attributes with a large number of values as root nodes. It means it prefers the attribute with a large number of distinct values.

C4.5, an improvement of ID3, uses Gain ratio which is a modification of Information gain that reduces its bias and is usually the best option. Gain ratio overcomes the problem with information gain by taking into account the number of branches that would result before making the split. It corrects information gain by taking the intrinsic information of a split into account.

Let us consider if we have a dataset that has users and their movie genre preferences based on variables like gender, group of age, rating, blah, blah. With the help of information gain, you split at ‘Gender’ (assuming it has the highest information gain) and now the variables ‘Group of Age’ and ‘Rating’ could be equally important and with the help of gain ratio, it will penalize a variable with more distinct values which will help us decide the split at the next level.

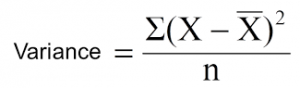


[Gain Ratio](https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147)

Where “before” is the dataset before the split, K is the number of subsets generated by the split, and (j, after) is subset j after the split.

### ****Reduction in Variance****

**Reduction in variance** is an algorithm used for continuous target variables (regression problems). This algorithm uses the standard formula of variance to choose the best split. The split with lower variance is selected as the criteria to split the population:



Above X-bar is the mean of the values, X is actual and n is the number of values.

**Steps to calculate Variance:**

1. Calculate variance for each node.
2. Calculate variance for each split as the weighted average of each node variance.

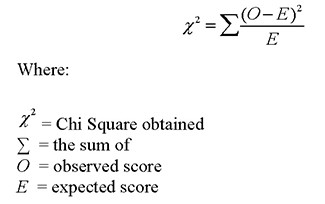
### ****Chi-Square****

The acronym CHAID stands for Chi-squared Automatic Interaction Detector. It is one of the oldest tree classification methods. It finds out the statistical significance between the differences between sub-nodes and parent node. We measure it by the sum of squares of standardized differences between observed and expected frequencies of the target variable.

It works with the categorical target variable “Success” or “Failure”. It can perform two or more splits. Higher the value of Chi-Square higher the statistical significance of differences between sub-node and Parent node.

It generates a tree called CHAID (Chi-square Automatic Interaction Detector).

Mathematically, Chi-squared is represented as:



**Steps to Calculate Chi-square for a split:**

1. Calculate Chi-square for an individual node by calculating the deviation for Success and Failure both
2. Calculated Chi-square of Split using Sum of all Chi-square of success and Failure of each node of the split

### ****How to avoid/counter Overfitting in Decision Trees?****

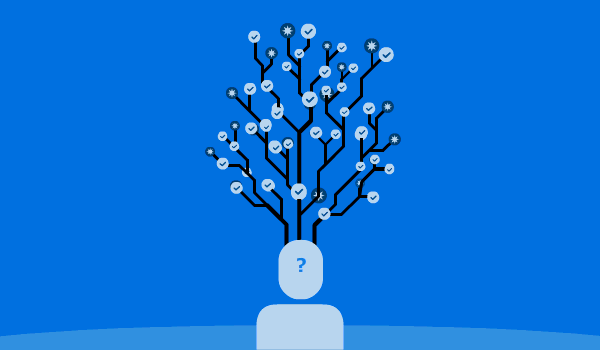
The common problem with Decision trees, especially having a table full of columns, they fit a lot. Sometimes it looks like the tree memorized the training data set. If there is no limit set on a decision tree, it will give you 100% accuracy on the training data set because in the worse case it will end up making 1 leaf for each observation. Thus this affects the accuracy when predicting samples that are not part of the training set.

Here are two ways to remove overfitting:

1. Pruning Decision Trees.
2. Random Forest

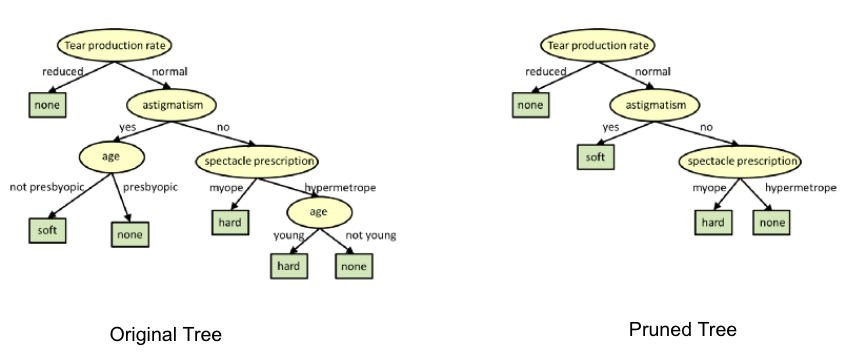
**Pruning Decision Trees**

The splitting process results in fully grown trees until the stopping criteria are reached. But, the fully grown tree is likely to overfit the data, leading to poor accuracy on unseen data.



[Pruning in action](https://gfycat.com/enchantedyellowishbarasinga)

In **pruning**, you trim off the branches of the tree, i.e., remove the decision nodes starting from the leaf node such that the overall accuracy is not disturbed. This is done by segregating the actual training set into two sets: training data set, D and validation data set, V. Prepare the decision tree using the segregated training data set, D. Then continue trimming the tree accordingly to optimize the accuracy of the validation data set, V.



[Pruning](https://www.cs.cmu.edu/~bhiksha/courses/10-601/decisiontrees/)

In the above diagram, the ‘Age’ attribute in the left-hand side of the tree has been pruned as it has more importance on the right-hand side of the tree, hence removing overfitting.

**Random Forest**

Random Forest is an example of ensemble learning, in which we combine multiple machine learning algorithms to obtain better predictive performance.

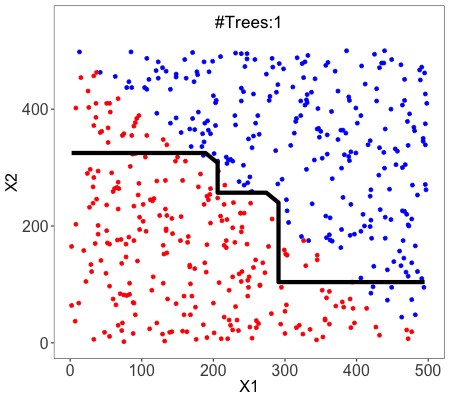
**Why the name “Random”?**

Two key concepts that give it the name random:

1. A random sampling of training data set when building trees.
2. Random subsets of features considered when splitting nodes.

A technique known as bagging is used to create an ensemble of trees where multiple training sets are generated with replacement.

In the bagging technique, a data set is divided into **N** samples using randomized sampling. Then, using a single learning algorithm a model is built on all samples. Later, the resultant predictions are combined using voting or averaging in parallel.



[Random Forest in action](https://towardsdatascience.com/why-random-forests-outperform-decision-trees-1b0f175a0b5)

### Which is better Linear or tree-based models?

Well, it depends on the kind of problem you are solving.

1. If the relationship between dependent & independent variables is well approximated by a linear model, linear regression will outperform the tree-based model.
2. If there is a high non-linearity & complex relationship between dependent & independent variables, a tree model will outperform a classical regression method.
3. If you need to build a model that is easy to explain to people, a decision tree model will always do better than a linear model. Decision tree models are even simpler to interpret than linear regression!

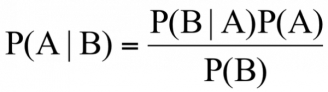
# Machine Learning Algorithms Explained - Naive Bayes Classifier

In our series **Machine Learning Algorithms Explained**, our goal is to give you a good sense of how the algorithms behind machine learning work, as well as the strengths and weaknesses of different methods. Each post in this series briefly explains a different algorithm – today, we’re going to talk about Naive Bayes Classifiers.

A Naive Bayes Classifier is a supervised machine-learning algorithm that uses the Bayes’ Theorem, which assumes that features are statistically independent. The theorem relies on the naive assumption that input variables are independent of each other, i.e. there is no way to know anything about other variables when given an additional variable. Regardless of this assumption, it has proven itself to be a classifier with good results.

# ****What Is the Bayes’ Theorem?****

Naive Bayes Classifiers rely on the Bayes’ Theorem, which is based on conditional probability or in simple terms, the likelihood that an event (A) will happen given that another event (B) has already happened. Essentially, the theorem allows a hypothesis to be updated each time new evidence is introduced. The equation below expresses Bayes’ Theorem in the language of probability:



Let’s explain what each of these terms means.

* “P” is the symbol to denote probability.
* P(A | B) = The probability of event A (hypothesis) occurring given that B (evidence) has occurred.
* P(B | A) = The probability of the event B (evidence) occurring given that A (hypothesis) has occurred.
* P(A) = The probability of event B (hypothesis) occurring.
* P(B) = The probability of event A (evidence) occurring.

# ****Drug Testing****

To better illustrate the usefulness of Bayes’ Theorem, let’s walk through an example. Suppose that a drug-testing method has the following statistical results:

* 0.5 percent of the population uses drugs.
* The test is positive for drug users 99 percent of the time (and therefore, it misses 1 percent of drug users).

|  |  |  |
| --- | --- | --- |
|  | Drug User (.05%) | Non-Drug user (99.5%) |
| Positive results | 99% | 1% |
| Negative results | 1% | 99% |

* The test is negative for a non-drug user 99 percent of the time (and positive for a non-drug user 1 percent of the time).

We would like to know the odds that a randomly selected individual who had a positive test result is actually a drug user, i.e. that the results are a true positive. In the language of probability, that looks something like this:

* P( User | Positive ) = The probability that an individual that had a positive test result is actually a drug user.

Following Bayes’ Theorem, we know that the probability is calculated as follows:

P( User | Positive ) = ( P( Positive | User ) \* P( User) ) / P(Positive)

Now, let’s calculate each of the terms on the right side of the equation.

* P( Positive | User ) = 0.99. This is the chance of returning true positive test results if you are a drug user.
* P( User) = 0.005. The percentage of the population that are drug users.
* P(Positive) = P(Positive | User) \* P(User) + P(Positive | Non-user) \* P(Non-user), which translates to: .99\*.005 + .01\*.995 = 0.0149. [This is the breakdown of how a positive result— a user testing positive(true positive) and a non-user testing positive (false positive)—can occur.]

Applying these figures to the original formula gives you:

P( User | Positive ) = (.99 \* .005 ) / .99 = 33%

What a surprising result! At first, the test seems highly accurate. However, upon further calculation, we see that it’s far more likely that even after testing positive a person is, in fact, not a drug-user.

# ****Naive Bayes Classifier****

Given a vector **x** of features, Naive Bayes calculates the probability that the vector belongs to each class.

P( Ck | x1, x2, … xn )

Thanks to Bayes’ Theorem, we know that it is equal to:

P( Ck | **x**) = (P(**x** | Ck) \*P(Ck) ) / P(**x**)

We know P(Ck) because of the class distribution in our data.

P(**x** | Ck) is equivalent to its joint probability P(Ck ,x1 , x2, …, xn). By the chain rule in probabilities we can expand it to:

P(Ck ,x1 , x2, …, xn) = P(x1 | x2, …, xn, Ck) \* P(x2, …, xn, Ck)

= P(x1 | x2, …, xn, Ck) \* P(x2, | x3,…, xn, Ck) \* P(x3,…, xn, Ck)

= ….

= P(x1 | x2, …, xn, Ck) \* P(x2, | x3,…, xn, Ck) \* P(xn-1,| xn, Ck)\* P(xn | Ck)\* P(Ck)

We now make a strong assumption on the conditional probabilities we just calculated. We assume that they are conditionally independent. In other words, knowing that { xi+1,…, xn } occurred doesn't affect the probability of xi occurring. Put more formally:

P(xi, | xi+1,…, xn, Ck) = P(xi, | Ck).

This means that: P(Ck ,x1 , x2, …, xn) =

To calculate each of the conditional class probabilities—P(xi | Ck )—we use a likelihood function to model the probability distribution. One of the most common is the Gaussian or normal distribution.

We need not calculate P(**x**) since it’s constant for each input. But you can calculate it in terms of the conditional class probabilities as illustrated below:

P(**x**) =https://blog.easysol.net/wp-content/uploads/2018/03/Image-3.jpg

To make a prediction, we choose the class that had the highest score.

# ****Naive Bayes Examples****

We are going to use the Iris dataset we used in the previous blogs to illustrate how Naive Bayes works. Let's suppose we measure an Iris Setosa and find the following measurements:

* Sepal length = 7 cm
* Sepal width = 3 cm
* Petal length = 5 cm
* Petal width = 2 cm

From our data we know that each class—Iris versicolor, Iris virginica, and Iris setosa—represents one-third of the data. Following Naive Bayes, we need to calculate the following conditional probabilities and categorize our measured flower with the class that has the highest probability. To do so, we are going to use the Gaussian measure of likelihood:

* P( Setosa | 7,3,5,2)
* P( Versicolor | 7,3,5,2)
* P( Virginica | 7,3,5,2)

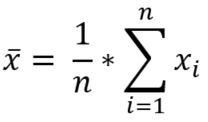
Let’s do one calculation with Versicolor:

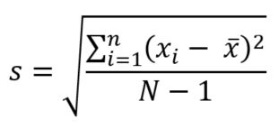
* P( Versicolor | 7,3,5,2) = (P( 7,3,5,2 | Versicolor ) \* P( Versicolor ) )/ P( 7,3,5,2)

Assuming independence and using the Gaussian distribution of conditional class probabilities, we can calculate the following:

* P( 7,3,5,2 | Versicolor ) = P(7 | Versicolor ) \* P(3 | Versicolor ) \* P(5 | Versicolor ) \* P(2 | Versicolor )

To calculate each of the conditional class probabilities, we have to find the average and standard deviation of each of the features operating under the assumption that they are Versicolor. The average and standard deviation are calculated as follows.

Average =

Standard deviation =

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sepal Length | Sepal Width | Petal Length | Petal Width |
| Average | 5.936 | 2.77 | 4.26 | 1.32 |
| Standard Deviation | 0.51 | 0.31 | 0.46 | 0.19 |

Plugging those values into a Gaussian distribution, we can calculate: ( N stands for the normal distribution)

P( 7,3,5,2 | Versicolor ) = P(7 | Versicolor ) \* P(3 | Versicolor ) \* P(5 | Versicolor ) \* P(2 | Versicolor )

P( 7,3,5,2 | Versicolor ) = N(7 | 5.936, 0.51) \* N(3 | 2.77, 0.31) \* N(5 | 4.26, 0.46) \* N(2 | 1.32, 0.19)

P( 7,3,5,2 | Versicolor ) = 0.089 \* 0.97 \* 0.24 \* 0.05

P( 7,3,5,2 | Versicolor ) = 0.001

P( Versicolor ) = 50/150

At last we can calculate: P( 7,3,5,2 | Versicolor ) 0.001\* 0.33 = .0003

From here, we would need to calculate the same process for Setosa and Virginica in order to determine in which class the original flower is most likely to belong.

# ****Summary****

Advantages of Naive Bayes Classifiers

* Simple model
* Fast
* Scalable
* Requires little data

Disadvantages of Naive Bayes Classifiers

* Assumes feature independence
* Must choose the likelihood function