**PROFESSIONAL TRAINING REPORT**

**at**

**Sathyabama Institute of Science and Technology**

**(Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering degree in Computer Science and Engineering

By

**VATTIKUTI MANIDEEP SITARAM**

**(Reg. No. 38110624)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A” by NAAC**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

**NOVEMBER 2021**

**SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY**

### (DEEMED TO BE UNIVERSITY)

**Accredited with Grade “A” by NAAC**

**(Established under Section 3 of UGC Act, 1956)**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI– 600119**

[**www.sathyabamauniversity.ac.in**](http://www.sathyabamauniversity.ac.in/)

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **V. MANIDEEP SITARAM (38110624)** who carried out the project entitled “**Parkinson Disease Prediction**” under my supervision from November 2019 to April 2020.

## Internal Guide Name: Mrs .S . L . Jany Shabu

**Head of the Department**

**Dr S. VIGNESHWARI,M.E,PH.D**



## Submitted for Viva voce Examination held on

**Internal Examiner External Examiner**

**DECLARATION**

I (name/s of the individual Candidate) V.Manideep Sitaram hereby declare that the Project Report entitled Prediction of Parkinson Disease done by me under the guidance of Dr. /Prof./ Mr./Ms..S . L Jany Shabu Department of Computer Science andEngineering at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering**.**

## DATE:

## V.Manideep Sitaram

**PLACE: SIGNATURE OF THE CANDIDATE**

**ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.S.Vigneshwari M.E., Ph.D., and Dr.L.Lakshmanan M.E., Ph.D.,** Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs S.L.Jany Shabu** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE



|  |  |  |
| --- | --- | --- |
|  | **TABLE OF CONTENTS** |  |
| **CHAPTER No.** | **TITLE** | **PAGE No** |
|  | ABSTRACT | i |
|  |  |  |
|  |  |  |
|  |
|  |  |
| **1** | **INTRODUCTION** | **1** |
|  |  |  |
|  |  |  |
| **2.** | **AIM AND SCOPE OF THE PRESENT INVESTIGATION** | **40** |
|  |  |  |
|  |  |  |
|  |  |  |

**3. EXPERIMENTAL OR MATERIALS AND METHODS; ALGORITHMS USED**

|  |  |  |
| --- | --- | --- |
|  | 3.1 Variables used | 42 |
|  | 3.2 Methods | 43 |
|  | 3.3 Modules | 44 |
| **4.** | **RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS** | **43** |
|  |  |  |
|  |  |  |
| **5.** | **SUMMARY AND CONCLUSIONS** | **47** |
|  |  |  |
|  | **APPENDIX** |  |
|  | 5.1.SCREENSHOTS | 49 |
|  | 5.2. SOURCE CODE | 56 |

**ABSTRACT**

Parkinson’s Disease is the second most prevalent neurodegenerative disorder after Alzheimer’s, affecting more than 10 million people worldwide. Parkinson’s is characterized primarily by the deterioration of motor and cognitive ability. There is no single test which can be administered for diagnosis. Instead, doctors must perform a careful clinical analysis of the patient’s medical history. Unfortunately, this method of diagnosis is highly inaccurate. A study from the National Institute of Neurological Disorders finds that early diagnosis (having symptoms for 5 years or less) is only 53% accurate. This is not much better than random guessing, but an early diagnosis is critical to effective treatment .Because of these difficulties, I investigate a machine learning approach to accurately diagnose Parkinson’s, using a dataset of various speech features (a non-invasive yet characteristic tool) from the University of Oxford.

**INTRODUCTION**

**1.1 WHAT IS PARKINSON DISEASE HOW TO PREDICT IT**

Parkinson’s disease is a neurological movement disorder. Common symptoms include tremor, slowness of movement, stiff muscles, unsteady walk and balance and coordination problems. There is no cure for the disease. Most patients can maintain a good quality of life with medications. In some patients, surgery can help improve symptoms.

## OVERVIEW

### What is Parkinson’s disease?

Parkinson’s disease is a nervous system disease that affects your ability to control movement. The disease usually starts out slowly and worsens over time. If you have Parkinson’s disease, you may shake, have muscle stiffness, and have trouble walking and maintaining your balance and coordination. As the disease worsens, you may have trouble talking, sleeping, have mental and memory problems, experience behavioral changes and have other symptoms.

### Who gets Parkinson’s disease?

About 50% more men than women get Parkinson’s disease. It is most commonly seen in persons 60 years of age and older. However, up to 10% of patients are diagnosed before age 50.

About 60,000 new cases of Parkinson’s disease are diagnosed in the United States each year.

### Is Parkinson’s disease inherited?

Scientists have discovered gene mutations that are associated with Parkinson’s disease.

There is some belief that some cases of early-onset Parkinson’s disease – disease starting before age 50 – may be inherited. Scientists identified a gene mutation in people with Parkinson’s disease whose brains contain Lewy bodies, which are clumps of the protein alpha-synuclein. Scientists are trying to understand the function of this protein and its relationship to genetic mutations that are sometimes seen in Parkinson’s disease and in people with a type of dementia called [Lewy body dementia](https://my.clevelandclinic.org/health/diseases/17815-lewy-body-dementia).

Several other gene mutations have been found to play a role in Parkinson’s disease. Mutations in these genes cause abnormal cell functioning, which affects the nerve cells’ ability to release dopamine and causes nerve cell death. Researchers are still trying to discover what causes these genes to mutate in order to understand how gene mutations influence the development of Parkinson’s disease.

Scientists think that about 10% to 15% of person’s with Parkinson’s disease may have a genetic mutation that predisposes them to development of the disease. There are also environmental factors involved that are not fully understood.

## SYMPTOMS AND CAUSES

### What causes Parkinson’s disease?

Parkinson’s disease occurs when nerve cells (neurons) in an area of the brain called the substantia nigra become impaired or die. These cells normally produce dopamine, a chemical (neurotransmitter) that helps the cells of the brain communicate (transmits signals, “messages,” between areas in the brain). When these nerve cells become impaired or die, they produce less dopamine. Dopamine is especially important for the operation of another area of the brain called the basal ganglia. This area of the brain is responsible for organizing the brain’s commands for body movement. The loss of dopamine causes the movement symptoms seen in people with Parkinson’s disease.

People with Parkinson’s disease also lose another neurotransmitter called norepinephrine. This chemical is needed for proper functioning of the sympathetic nervous system. This system controls some of the body’s autonomic functions such as digestion, heart rate, blood pressure and breathing. Loss of norepinephrine causes some of the non-movement-related symptoms of Parkinson’s disease.

Scientists aren’t sure what causes the neurons that produce these neurotransmitter chemicals to die.

### What are the symptoms of Parkinson’s disease?

Symptoms of Parkinson’s disease and the rate of decline vary widely from person to person. The most common symptoms include:

* **Tremor:** Shaking begins in your hands and arms. It can also occur in your jaw or foot. In the early stages of the disease, usually only one side of your body or one limb is affected. As the disease progresses, tremor may become more wide spread. It worsens with stress. Tremor often disappears during sleep and when your arm or leg is being moved.
* **Slowness of movement (bradykinesia):** This is the slowing down of movement and is caused by your brain’s slowness in transmitting the necessary instructions to the appropriate parts of the body. This symptom is unpredictable and can be quickly disabling. One moment you may be moving easily, the next you may need help moving at all and finishing tasks such as getting dressed, bathing or getting out of a chair. You may even drag your feet as you walk.
* **Rigid muscles/stiff limbs:** Rigidity is the inability of your muscles to relax normally. This rigidity is caused by uncontrolled tensing of your muscles and results in you not being able to move about freely. You may experience aches or pains in the affected muscles and your range of motion may be limited.
* **Unsteady walk and balance and coordination problems:** You may develop a forward lean that makes you more likely to fall when bumped. You may take short shuffling steps, have difficulty starting to walk and difficulty stopping and not swing your arms naturally as you walk. You may feel like your feet are stuck to the floor when trying to take a step.
* **Muscle twisting, spasms or cramps (**[**dystonia**](https://my.clevelandclinic.org/health/articles/6006-dystonias)**).** You may experience a painful cramp in your foot or curled and clenched toes. Dystonia can occur in other body parts.
* **Stooped posture**. You have a “hunched over” posture.

Other symptoms include:

* Decreased facial expressions: You may not smile or blink as often as the disease worsens; your face lacks expression.
* Speech/vocal changes: Speech may be quick, become slurred or be soft in tone. You may hesitate before speaking. The pitch of your voice may become unchanged (monotone).
* Handwriting changes: You handwriting may become smaller and more difficult to read.
* Depression and anxiety.
* Chewing and swallowing problems, drooling.
* Urinary problems.
* Mental “thinking” difficulties/memory problems.
* Hallucinations/delusions.
* [Constipation.](https://my.clevelandclinic.org/health/articles/constipation)
* Skin problems, such as [dandruff.](https://my.clevelandclinic.org/health/articles/seborrheic-dermatitis)
* Loss of smell.
* [Sleeping disturbances](https://my.clevelandclinic.org/health/articles/9366-sleep-problems-with-parkinsons-disease) including disrupted sleep, acting out your dreams, and [restless leg syndrome](https://my.clevelandclinic.org/health/diseases/9497-restless-legs-syndrome).
* Pain, lack of interest (apathy), fatigue, change in weight, vision changes.
* Low blood pressure.

### What are the different stages of Parkinson’s disease?

Each person with Parkinson’s disease experiences symptoms in in their own unique way. Not everyone experiences all symptoms of Parkinson’s disease. You may not experience symptoms in the same order as others. Some people may have mild symptoms; others may have intense symptoms. How quickly symptoms worsen also varies from individual to individual and is difficult to impossible to predict at the outset.

In general, the disease progresses from early stage to mid-stage to mid-late-stage to advanced stage. This is what typically occurs during each of these stages:

**Early stage**

Early symptoms of Parkinson’s disease are usually mild and typically occur slowly and do not interfere with daily activities. Sometimes early symptoms are not easy to detect or you may think early symptoms are simply normal signs of aging. You may have fatigue or a general sense of uneasiness. You may feel a slight tremor or have difficulty standing.

Often, a family member or friend notices some of the subtle signs before you do. They may notice things like body stiffness or lack of normal movement (no arm swing when walking) slow or small handwriting, lack of expression in your face, or difficulty getting out of a chair.

**Mid stage**

Symptoms start getting worse. Tremor, muscle stiffness and movement problems may now affect both sides of the body. Balance problems and falls are becoming more common. You may still be fully independent but daily tasks of everyday living, such as bathing and dressing, are becoming more difficult to do and take longer to complete.

**Mid-late stage**

Standing and walking are becoming more difficult and may require assistance with a walker. You may need full time help to continue to live at home.

**Advanced stage**

You now require a wheelchair to get around or are bedridden. You may experience hallucinations or delusions. You now require full-time nursing care.

## DIAGNOSIS AND TESTS

### How is Parkinson’s disease diagnosed?

Diagnosing Parkinson’s disease is sometimes difficult, since early symptoms can mimic other disorders and there are no specific blood or other laboratory tests to diagnose the disease. Imaging tests, such as [CT (computed tomography)](https://my.clevelandclinic.org/health/articles/computed-tomography-ct-scan) or [MRI (magnetic resonance imaging)](https://my.clevelandclinic.org/health/articles/magnetic-resonance-imaging-mri) scans, may be used to rule out other disorders that cause similar symptoms.

To diagnose Parkinson’s disease, you will be asked about your medical history and family history of neurologic disorders as well as your current symptoms, medications and possible exposure to toxins. Your doctor will look for signs of tremor and muscle rigidity, watch you walk, check your posture and coordination and look for slowness of movement.

If you think you may have Parkinson’s disease, you should probably see a neurologist, preferably a movement disorders-trained neurologist. The treatment decisions made early in the illness can affect the long-term success of the treatment.

## MANAGEMENT AND TREATMENT

### How is Parkinson’s disease treated?

There is no cure for Parkinson’s disease. However, medications and other treatments can help relieve some of your symptoms. Exercise can help your Parkinson’s symptoms significantly. In addition, physical therapy, occupational therapy and [speech-language therapy](https://my.clevelandclinic.org/health/diseases/9392-speech-therapy-for-parkinsons-disease) can help with walking and balance problems, eating and swallowing challenges and speech problems. Surgery is an option for some patients.

### What medications are used to treat Parkinson’s disease?

Medications are the main treatment method for patients with Parkinson’s disease. Your doctor will work closely with you to develop a treatment plan best suited for you based on the severity of your disease at the time of diagnosis, side effects of the drug class and success or failure of symptom control of the medications you try.

Medications combat Parkinson’s disease by:

* Helping nerve cells in the brain make dopamine.
* Mimicking the effects of dopamine in the brain.
* Blocking an enzyme that breaks down dopamine in the brain.
* Reducing some specific symptoms of Parkinson’s disease.

**Levodopa:** Levodopa is a main treatment for the slowness of movement, tremor, and stiffness symptoms of Parkinson’s disease. Nerve cells use levodopa to make dopamine, which replenishes the low amount found in the brain of persons with Parkinson’s disease. Levodopa is usually taken with carbidopa (Sinemet®) to allow more levodopa to reach the brain and to prevent or reduce the nausea and vomiting, low blood pressure and other side effects of levodopa. Sinemet® is available in an immediate release formula and a long-acting, controlled release formula. Rytary® is a newer version of levodopa/carbidopa that is a longer-acting capsule. The newest addition is Inbrija®, which is inhaled levodopa. It is used by people already taking regular carbidopa/levodopa for when they have off episodes (discussed below).

As people have Parkinson’s for a longer amount of time, the effects of their levodopa doses don't last as long as they did before, resulting in their symptoms (tremor, muscle rigidity, slowness) worsening before they are due to take their next dose. This is called ‘wearing off.’ They may also notice involuntary, fluid, dancing or fidgeting-like movements of their body called dyskinesias. These movements can indicate the levodopa dose is too high. These ups and downs of the effects of levodopa are called motor fluctuations and are often improve with adjustment of the medication by the neurologist.

**Dopamine agonists:** These drugs mimic the effects of dopamine in your brain. They are not as effective as levodopa in controlling slow muscle movement and muscle rigidity. Your doctor may try these medications first and add levodopa if your symptoms are not well controlled depending on severity of your symptoms and your age.

Newer dopamine medications include ropinirole (Requip®) and pramipexole (Mirapex®). Rotigotine (Neupro®) is given as a patch. Apomorphine (Apokyn®) is a short-acting injectable medication.

Side effects of dopamine agonists include nausea, vomiting, dizziness, lightheadedness, sleeping problems, leg swelling, confusion, hallucinations and compulsive behavior (such as excessive gambling, buying, eating, or sex). Some of these side effects are more likely to occur in people over 70 years old.

**Catechol O-methyltransferase (COMT) inhibitors:** These drugs block an enzyme that breaks down dopamine in your brain. These drugs are taken with levodopa and slow your body’s ability to get rid of levodopa, so it lasts longer and is more reliable. Entacapone (Comtan®) and tolcapone (Tasmar®) are examples of COMT inhibitors. Opicapone (Ongentys®) is the newest medication in this class, receiving FDA approval in April 2020. Because these drugs increase the effectiveness of levodopa, they may also increase its side effects, including involuntary movements (dyskinesia). Tolcapone is rarely prescribed because it can damage the liver and requires close monitoring to prevent liver failure.

**MAO B inhibitors.** These drugs block a particular brain enzyme – monoamine oxidase B (MAO B) – that breaks down dopamine in your brain. This allows dopamine to have longer lasting effects on the brain. Examples of MAO B inhibitors include selegiline (Eldepryl®, Zelapar®), rasagiline (Azilect®) and safinamide (Xadago®). Side effects of these drugs include nausea and insomnia. Giving carbidopa-levodopa with an MAO B inhibitor increases the chance of hallucinations and dyskinesia. MAO B inhibitors are not prescribed if you are taking certain antidepressants or narcotic medications. Your doctor will review all your current medications and make the best treatment choice for you.

**Anticholinergics.** These drugs help reduce tremor and muscle stiffness. Examples include benztropine (Cogentin®) and trihexyphenidyl (Artane®). These are the oldest class of drugs to treat Parkinson’s disease. Side effects include blurred vision, constipation, dry mouth and urine retention. Persons over age 70 who are prone to confusion and hallucinations or have memory impairment should not take anticholinergics. Because of the high rate of side effects these medications are less commonly used.

**Amantadine.** Amantadine (Symmetrel®), first developed as an antiviral agent, is useful in reducing the involuntary movements (dyskinesia) caused by levodopa medication. There are two extended-release forms of the drug, Gocovri®, and Osmolex ER®. Side effects include confusion and memory problems.

**Istradefylline.**Istradefylline (Nourianz®) is an adenosine A2A receptor antagonist. It is used for people taking carbidopa-levodopa but experiencing off symptoms. Like the other drugs that act to increase the effectiveness of levodopa, they may also increase its side effects, including involuntary movements (dyskinesia) and hallucinations.

### What are the surgical treatments for Parkinson’s disease?

Most patients with Parkinson’s disease can maintain a good quality of life with medications. However, as the disease worsens, medications may no longer be effective in some patients. In these patients, the effectiveness of medications becomes unpredictable – reducing symptoms during “on” periods and no longer controlling symptoms during “off” periods, which usually occur when the medication is wearing off and just before the next dose is to be taken. Sometimes these variations can be managed with changes in medications. However, sometimes they can’t. Based on the type and severity of your symptoms, the failure of adjustments in your medications, the decline in your quality of life and your overall health, your doctor may discuss some of the available surgical options.

* [**Deep brain stimulation**](https://my.clevelandclinic.org/health/treatments/4080-deep-brain-stimulation-for-parkinsons-disease-patients) (DBS) involves implanting electrodes in the brain, which deliver electrical impulses that block or change the abnormal activity that cause symptoms. DBS can treat most of the major movement symptoms of Parkinson’s disease such as tremor, slowness of movement (bradykinesia) and stiffness (rigidity). It does not improve memory, hallucinations, depression, and the other non-movement symptoms of Parkinson’s disease. Only patients whose symptoms are not controlled despite medication trials and who meet other strict criteria may be candidates for DBS. Your doctor will discuss if this is the right treatment for you.
* **Carbidopa-levodopa infusion** involves the surgical placement of a feeding tube into the small intestine. A gel form of the medication carbidopa-levodopa (Duopa®) is delivered through this tube. This method of continuous infusion of the drug keeps a stable dosage in the body. This helps patients who have had variation in their response to the oral form of carbidopa-levodopa but are still benefitting from the combination drug.
* **Pallidotomy** involves destroying a small portion of a part of the brain that controls movement (the globus pallidus). Pallidotomy help reduce involuntary movements (dyskinesias), muscle stiffness and tremor.
* **Thalamotomy** involves destroying a small part of the thalamus. This may help a small number of patients who have severe tremors of their arm or hand.

## PREVENTION

### Can Parkinson’s disease be prevented?

Unfortunately, no. Parkinson’s disease is long-term disease that worsens over time. Although there is no way to prevent or cure the disease (at this current moment in time), medications may significantly relieve your symptoms. In some patients – especially those with later-stage disease, surgery to improve symptoms may be an option.

## OUTLOOK / PROGNOSIS

### What is the outlook for persons with Parkinson’s disease?

Although there is no cure or absolute evidence of ways to prevent Parkinson’s disease, scientists are working hard to learn more about the disease and find innovative ways to better manage it, prevent it from progressing and ultimately curing it.

Currently, you and your healthcare team’s efforts are focused on medical management of your symptoms along with general health and lifestyle improvement recommendations (exercise, healthy eating, improved sleep). By identifying individual symptoms and adjusting the course of action based on changes in symptoms, most people with Parkinson’s disease can live fulfilling lives.

The future is hopeful. Some of the research underway includes:

* Using stem cells (from either bone marrow or embryos) to produce new neurons, which would produce dopamine.
* Producing a dopamine-producing enzyme that is delivered to a gene in the brain that controls movement.
* Using a naturally occurring human protein – glial cell-line derived neurotrophic factor, GDNF – to protect dopamine-releasing nerve cells.

Many other investigations are underway too. Much has been learned, much progress has been made and additional discoveries are likely to come.

## LIVING WITH

### What lifestyle changes can I make to ease Parkinson’s symptoms?

[**Exercise:**](https://my.clevelandclinic.org/health/articles/9200-exercise-for-people-with-parkinsons-disease) Exercise helps improve muscle strength, balance, coordination, flexibility, and tremor. It is also strongly believed to improve memory, thinking and reduce the risk of falls and decrease anxiety and depression. One study in persons with Parkinson’s disease showed that 2.5 hours of exercise per week resulted in improved ability to move and a slower decline in quality of life compared to those who didn’t exercise or didn’t start until later in the course of their disease. Some exercises to consider include strengthening or resistance training, stretching exercises or aerobics (running, walking, dancing). All types of exercise are helpful.

**Eat a healthy, balanced diet:** This is not only good for your general health but can ease some of the non-movement related symptoms of Parkinson’s, such as constipation. [Eating foods high in fiber](https://my.clevelandclinic.org/health/articles/14400-improving-your-health-with-fiber) in particular can relieve constipation. The [Mediterranean diet](https://my.clevelandclinic.org/health/articles/16037-mediterranean-diet) is one example of a healthy diet.

**Preventing falls and maintaining balance:** Falls are a frequent complication of Parkinson's. While you can do many things to reduce your risk of falling, the two most important are: 1) to work with your doctor to ensure that your treatments — whether medicines or deep brain stimulation — are optimal; and 2) to consult with a physical therapist who can assess your walking and balance. The physical therapist is the expert when it comes to recommending assistive devices or exercise to improve safety and preventing falls.

[**Improve the quality of your sleep.**](https://my.clevelandclinic.org/health/articles/9366-sleep-problems-with-parkinsons-disease)

### How do I prevent falls from common hazards?

* **Floors:**Remove all loose wires, cords, and throw rugs. Minimize clutter. Make sure rugs are anchored and smooth. Keep furniture in its usual place.
* **Bathroom:**Install grab bars and non-skid tape in the tub or shower. Use non-skid bath mats on the floor or install wall-to-wall carpeting.
* **Lighting:**Make sure halls, stairways, and entrances are well-lit. Install a night light in your bathroom or hallway and staircase. Turn lights on if you get up in the middle of the night. Make sure lamps or light switches are within reach of the bed if you have to get up during the night.
* **Kitchen:**Install non-skid rubber mats near the sink and stove. Clean spills immediately.
* **Stairs:**Make sure treads, rails, and rugs are secure. Install a rail on both sides of the stairs. If stairs are a threat, it might be helpful to arrange most of your activities on the lower level to reduce the number of times you must climb the stairs.
* **Entrances and doorways:**Install metal handles on the walls adjacent to the doorknobs of all doors to make it more secure as you travel through the doorway.

### What are some tips to help me maintain balance?

* Keep at least one hand free at all times. Try using a backpack or fanny pack to hold things rather than carrying them in your hands. Never carry objects in both hands when walking as this interferes with keeping your balance.
* Attempt to swing both arms from front to back while walking. This might require a conscious effort if Parkinson's disease has diminished your movement. It will, however, help you to maintain balance and posture, and reduce falls.
* Consciously lift your feet off of the ground when walking. Shuffling and dragging of the feet is a common culprit in losing your balance.
* When trying to navigate turns, use a "U" technique of facing forward and making a wide turn, rather than pivoting sharply.
* Try to stand with your feet shoulder-length apart. When your feet are close together for any length of time, you increase your risk of losing your balance and falling.
* Do one thing at a time. Don't try to walk and accomplish another task, such as reading or looking around. The decrease in your automatic reflexes complicates motor function, so the less distraction, the better.
* Do not wear rubber or gripping soled shoes--they might "catch" on the floor and cause tripping.
* Move slowly when changing positions. Use deliberate, concentrated movements and, if needed, use a grab bar or walking aid. Count 15 seconds between each movement. For example, when rising from a seated position, wait 15 seconds after standing to begin walking.
* If you become "frozen," visualize stepping over an imaginary object, or have someone place his or her foot in front of yours to step over. Try not to have a caregiver or family member "pull" you--this might throw you off balance and even prolong the episode.
* If balance is a continuous problem, you might want to consider a walking aid such as a cane, walking stick, or walker. Once you've mastered walking with help, you might be ready to try it on your own again.

## RESOURCES

Living with Parkinson’s disease can be a frustrating experience. It’s normal to feel angry, depressed and anxious. You and your family members might find it helpful to reach out to others who have this disease – to share your knowledge and insights, experiences and tips for living. You may want to check out local support groups of these Parkinson’s organizations:

* [American Parkinson Disease Association.](https://www.apdaparkinson.org/community/)
* Parkinson’s Foundation. [Find a Chapter.](https://www.parkinson.org/get-involved/local-resources)
* Parkinson & Movement Disorder (PDM) Alliance..

### Classification

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.

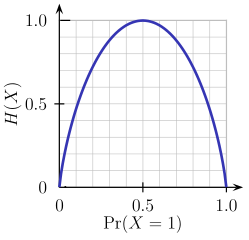
For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some criteria like :

**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****

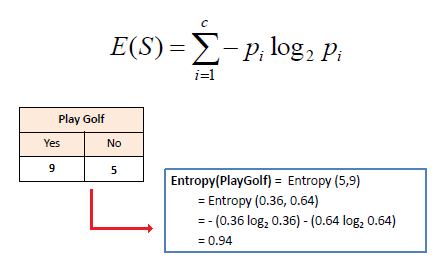
Entropy is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information. Flipping a coin is an example of an action that provides information that is random.



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.

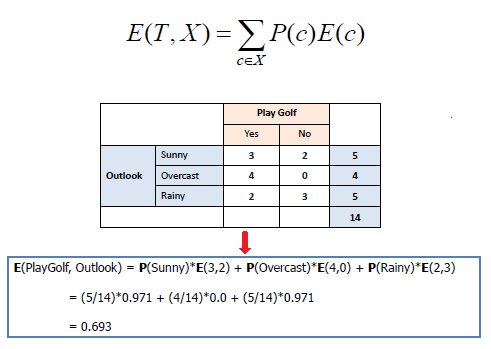
**ID3 follows the rule — A branch with an entropy of zero is a leaf node and A brach with entropy more than zero needs further splitting.**

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.**

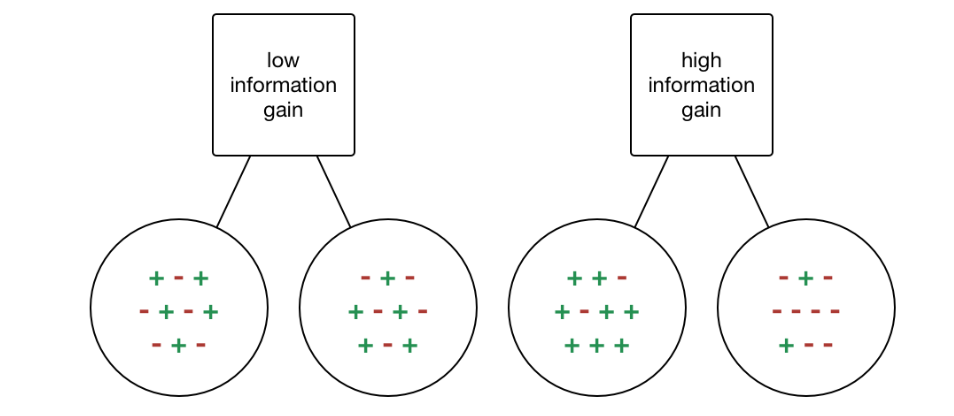
Mathematically Entropy for multiple attributes is represented as:



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

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

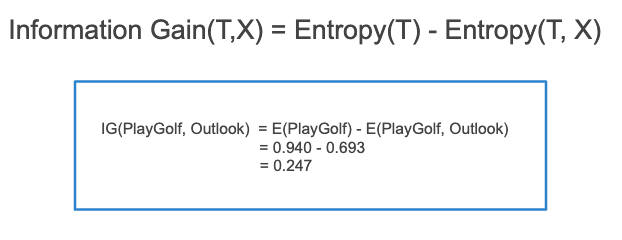
Information gain or **IG**is a statistical property that measures how well a given attribute separates the training examples according to their target classification. Constructing a decision tree is all about finding an attribute that returns the highest information gain and the smallest entropy.



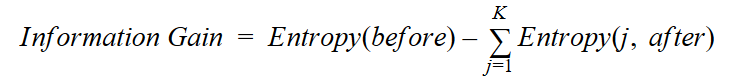
[Information Gain](https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a?gi=a8ffb5170258)

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:

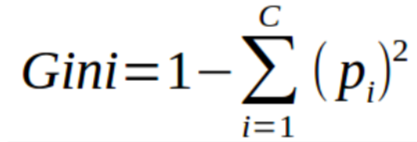


[Information Gain](https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147)

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 partitions 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 value of Gini index implies higher inequality, higher heterogeneity.

**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.

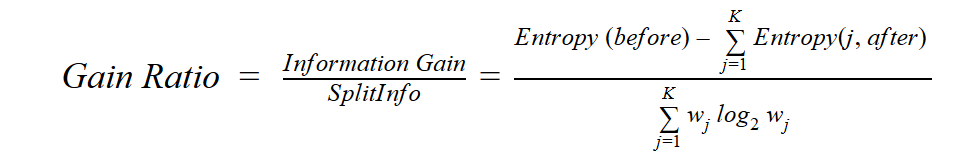
CART (Classification and Regression Tree) uses the Gini index method to create split points.

### 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.

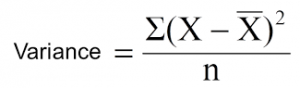


[Gain Ratio](https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147)

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.

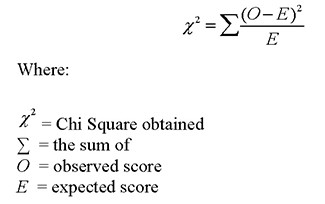
### ****Chi-Square****

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?****

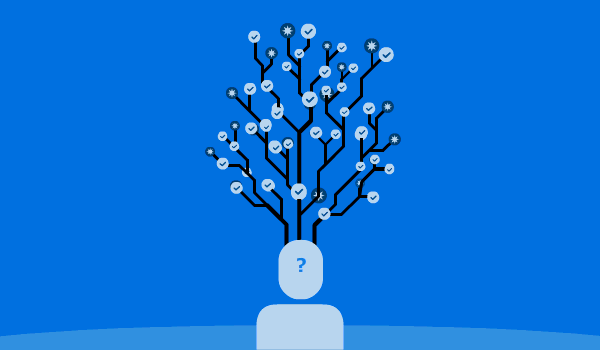
The common problem with Decision trees, especially having a table full of columns, they fit a lot. Sometimes it looks like the tree memorized the training data set. If there is no limit set on a decision tree, it will give you 100% accuracy on the training data set because in the worse case it will end 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

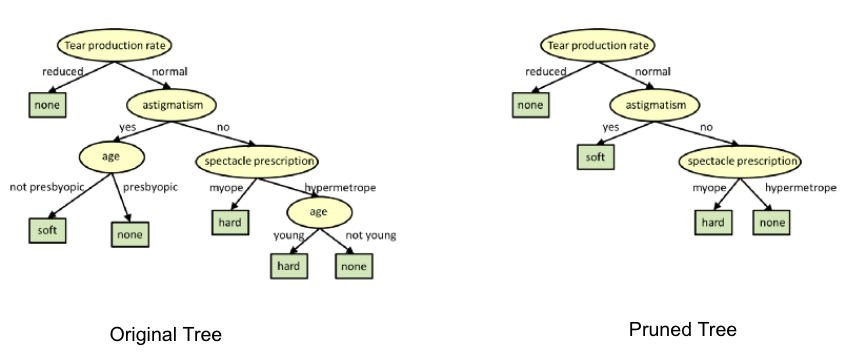
**Pruning Decision Trees**

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.

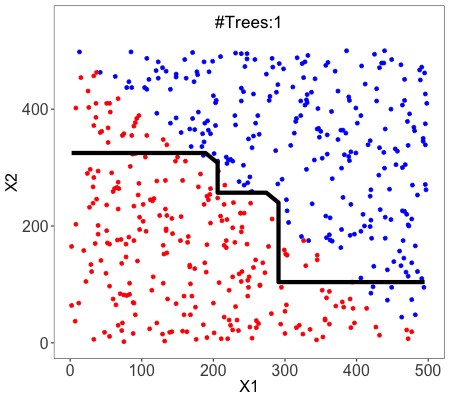
**Why the name “Random”?**

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!

### Decision Tree Classifier Building in Scikit-learn

The dataset that we have is a supermarket data which can be downloaded from [here](https://drive.google.com/open?id=1x1KglkvJxNn8C8kzeV96YePFnCUzXhBS).  
Load all the basic libraries.

import numpy as np

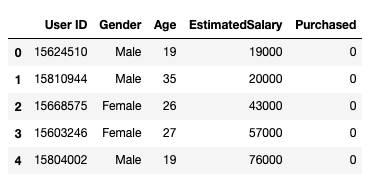
import matplotlib.pyplot as plt

import pandas as pd

Load the dataset. It consists of 5 features, UserID, Gender, Age, EstimatedSalary and Purchased.

data = pd.read\_csv('/Users/ML/DecisionTree/Social.csv')

data.head()



Dataset

We will take only Age and EstimatedSalary as our independent variables X because of other features like Gender and User ID are irrelevant and have no effect on the purchasing capacity of a person. Purchased is our dependent variable y.

feature\_cols = ['Age','EstimatedSalary' ]X = data.iloc[:,[2,3]].values

y = data.iloc[:,4].values

The next step is to split the dataset into training and test.

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X,y,test\_size = 0.25, random\_state= 0)

Perform feature scaling

#feature scaling

from sklearn.preprocessing import StandardScaler

sc\_X = StandardScaler()

X\_train = sc\_X.fit\_transform(X\_train)

X\_test = sc\_X.transform(X\_test)

Fit the model in the Decision Tree classifier.

from sklearn.tree import DecisionTreeClassifier

classifier = DecisionTreeClassifier()

classifier = classifier.fit(X\_train,y\_train)

Make predictions and check accuracy.

#prediction

y\_pred = classifier.predict(X\_test)#Accuracy

from sklearn import metricsprint('Accuracy Score:', metrics.accuracy\_score(y\_test,y\_pred))

The decision tree classifier gave an accuracy of 91%.

Confusion Matrix

from sklearn.metrics import confusion\_matrix

cm = confusion\_matrix(y\_test, y\_pred)Output:

array([[64, 4],

[ 2, 30]])

It means 6 observations have been classified as false.

**Let us first visualize the model prediction results.**

from matplotlib.colors import ListedColormap

X\_set, y\_set = X\_test, y\_test

X1, X2 = np.meshgrid(np.arange(start = X\_set[:,0].min()-1, stop= X\_set[:,0].max()+1, step = 0.01),np.arange(start = X\_set[:,1].min()-1, stop= X\_set[:,1].max()+1, step = 0.01))

plt.contourf(X1,X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape), alpha=0.75, cmap = ListedColormap(("red","green")))plt.xlim(X1.min(), X1.max())

plt.ylim(X2.min(), X2.max())for i,j in enumerate(np.unique(y\_set)):

plt.scatter(X\_set[y\_set==j,0],X\_set[y\_set==j,1], c = ListedColormap(("red","green"))(i),label = j)

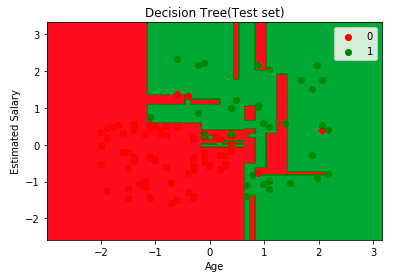
plt.title("Decision Tree(Test set)")

plt.xlabel("Age")

plt.ylabel("Estimated Salary")

plt.legend()

plt.show()



**Let us also visualize the tree:**

You can use Scikit-learn’s export\_graphviz function to display the tree within a Jupyter notebook. For plotting trees, you also need to install Graphviz and pydotplus.

conda install python-graphviz  
pip install pydotplus

export\_graphviz function converts decision tree classifier into dot file and pydotplus convert this dot file to png or displayable form on Jupyter.

from sklearn.tree import export\_graphviz

from sklearn.externals.six import StringIO

from IPython.display import Image

import pydotplusdot\_data = StringIO()

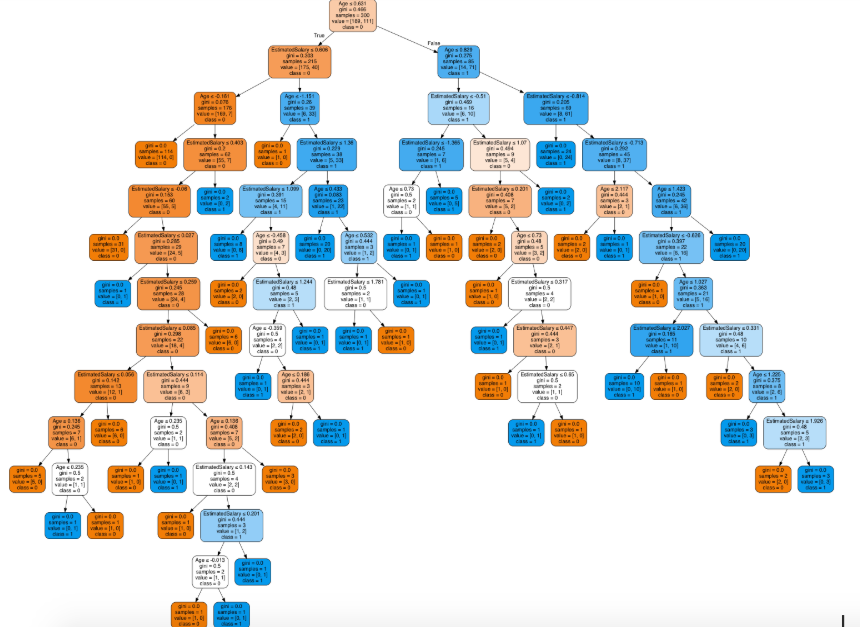
export\_graphviz(classifier, out\_file=dot\_data,

filled=True, rounded=True,

special\_characters=True,feature\_names = feature\_cols,class\_names=['0','1'])

graph = pydotplus.graph\_from\_dot\_data(dot\_data.getvalue())

Image(graph.create\_png())



Decision Tree.

In the decision tree chart, each internal node has a decision rule that splits the data. Gini referred to as the Gini ratio, which measures the impurity of the node. You can say a node is pure when all of its records belong to the same class, such nodes known as the leaf node.

Here, the resultant tree is unpruned. This unpruned tree is unexplainable and not easy to understand. In the next section, let’s optimize it by pruning.

**Optimizing the Decision Tree Classifier**

**criterion**: optional (default=”gini”) or Choose attribute selection measure: This parameter allows us to use the different-different attribute selection measure. Supported criteria are “gini” for the Gini index and “entropy” for the information gain.

**splitter**: string, optional (default=”best”) or Split Strategy: This parameter allows us to choose the split strategy. Supported strategies are “best” to choose the best split and “random” to choose the best random split.

**max\_depth**: int or None, optional (default=None) or Maximum Depth of a Tree: The maximum depth of the tree. If None, then nodes are expanded until all the leaves contain less than min\_samples\_split samples. The higher value of maximum depth causes overfitting, and a lower value causes underfitting (Source).

In Scikit-learn, optimization of decision tree classifier performed by only pre-pruning. The maximum depth of the tree can be used as a control variable for pre-pruning.

# Create Decision Tree classifer object

classifier = DecisionTreeClassifier(criterion="entropy", max\_depth=3)# Train Decision Tree Classifer

classifier = classifier.fit(X\_train,y\_train)#Predict the response for test dataset

y\_pred = classifier.predict(X\_test)# Model Accuracy, how often is the classifier correct?

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))

Well, the classification rate increased to 94%, which is better accuracy than the previous model.

Now let us again visualize the pruned Decision tree after optimization.

dot\_data = StringIO()

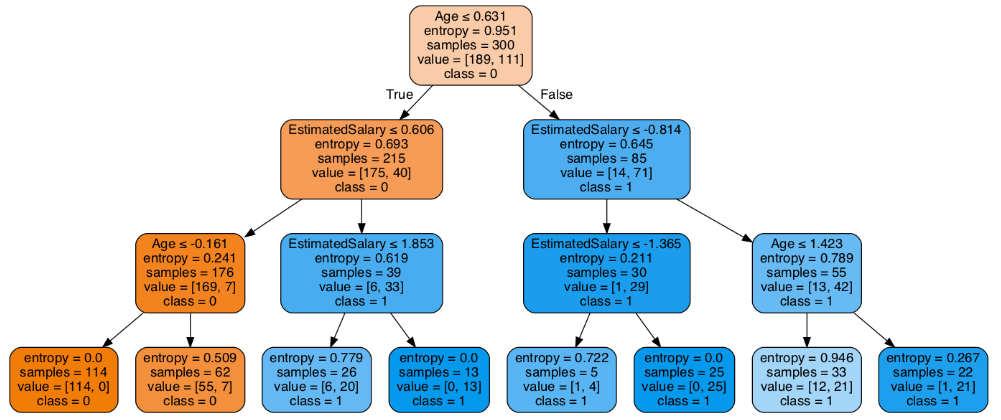
export\_graphviz(classifier, out\_file=dot\_data,

filled=True, rounded=True,

special\_characters=True, feature\_names = feature\_cols,class\_names=['0','1'])

graph = pydotplus.graph\_from\_dot\_data(dot\_data.getvalue())

Image(graph.create\_png())



Decision Tree after pruning

This pruned model is less complex, explainable, and easy to understand than the previous decision tree model plot.

### Conclusion

In this article, we have covered a lot of details about Decision Tree; It’s working, attribute selection measures such as Information Gain, Gain Ratio, and Gini Index, decision tree model building, visualization and evaluation on supermarket dataset using Python Scikit-learn package and optimizing Decision Tree performance using parameter tuning.

Well, that’s all for this article hope you guys have enjoyed reading it, feel free to share your comments/thoughts/feedback in the comment section.

2.**AIM AND SCOPE OF THE PRESENT**

**INVESTIGATION**

**2.1 AIM OF THE PROGRAM**

The Aim of the Program is to predict weather a person has Parkinson disease has or not by using classification model which takes the input as the feature of the model

**2.2 SCOPE OF THE INVESTIGATION**

1. The following program requires sklearn libraries to be imported which are already inbuilt in the jupyter notebook
2. The Goal here is to check weather a person has Parkinson disease or not
3. Read the dataset from the csv file which is downloaded from the Kaggle.com
4. Now do the data preprocessing , data scaling, data analysis on the data
5. Now split the data into training and testing data and fit the training data into the classification model.
6. After training of the model we use various classification metrics to find how well does our model perfoms

**3.EXPERIMENT ,MATERIALS OR METHODS AND ALOGORITHMS USED**

**3a. VARIABLES USED**

1.Name

2.MDVP:Fo(Hz)

3.MDVP:Fhi(Hz)

4.MDVP:Flo(Hz)

5.MDVP:Jitter(%)

6.MDVP:Jitter(Abs)

7.MDVP:RAP

8.MDVP:PPQJitter:DDP

9.MDVP:Shimmer

10.MDVP:Shimmer(dB)

11.Shimmer:APQ3

12.Shimmer:APQ5

13.MDVP:APQ

14.Shimmer:DDA

15.NHR

16.HNR

17.status

18.RPDE

19.DFA

20.spread1

21.spread2

22.D2

23.PPE

**3b METHODS USED**

1.LogisticRegression():

An classification model uses sigmoid function and linear line equation to predict the variable

2. DecisionTreeClassifier():

An classification model uses information gain ,entropy , gini impurity to build a tree and predict the output using that tree

3. KNeighborsClassifier():

An classification model uses eucleadian distance to find the distance between the output to be predicted and odd number of nearest points using which class has majority it predicts the output.

**3c MODULES USED**

***NumPy-***

NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

In Python we have lists that serve the purpose of arrays, but they are slow to process.

NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy.

***Pandas-***

Pandas is a [Python](https://www.python.org/) package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis/manipulation tool available in any language. It is already well on its way toward this goal.

pandas is well suited for many different kinds of data:

* Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
* Ordered and un ordered (not necessarily fixed-frequency) time series data.
* Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
* Any other form of observational / statistical data sets. The data need not be labeled at all to be placed into a pandas data structure

The two primary data structures of pandas, [Series](#pandas.Series) (1-dimensional) and [Data Frame](#pandas.DataFrame) (2-dimensional), handle the vast majority of typical use cases in finance, statistics, social science, and many areas of engineering. For R users, [Data Frame](#pandas.DataFrame) provides everything that R’s data. Frame provides and much more. pandas is built on top of [NumPy](https://www.numpy.org/) and is intended to integrate well within a scientific computing environment with many other 3rd party libraries.

Here are just a few of the things that pandas does well:

* Easy handling of missing data (represented as NaN) in floating point as well as non-floating-point data
* Size mutability: columns can be inserted and deleted from Data Frame and higher dimensional objects
* Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, Data Frame, etc. automatically align the data for you in computations
* Powerful, flexible group by functionality to perform split-apply-combine operations on data sets, for both aggregating and transforming data
* Make it easy to convert ragged, differently-indexed data in other Python and NumPy data structures into Data Frame objects
* Intelligent label-based slicing, fancy indexing, and sub setting of large data sets
* Intuitive merging and joining data sets.

***Scikit-learn-***

Scikit-learn (formerly skits. learn and also known as sklearn) is a [free software](https://en.wikipedia.org/wiki/Free_software) [machine learning](https://en.wikipedia.org/wiki/Machine_learning) [library](https://en.wikipedia.org/wiki/Library_(computing)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) [programming language](https://en.wikipedia.org/wiki/Programming_language).[[3]](#cite_note-jmlr-3) It features various [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support vector machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [*k*-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN](https://en.wikipedia.org/wiki/DBSCAN), and is designed to inter operate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy) and [SciPy](https://en.wikipedia.org/wiki/SciPy).

Scikit-learn is largely written in Python, and uses [numpy](https://en.wikipedia.org/wiki/Numpy) extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in [Cpython](https://en.wikipedia.org/wiki/Cython) to improve performance. Support vector machines are implemented by a Cpython wrapper around [LIBSVM](https://en.wikipedia.org/wiki/LIBSVM); logistic regression and linear support vector machines by a similar wrapper around [LIBLINEAR](https://en.wikipedia.org/wiki/LIBLINEAR). In such cases, extending these methods with Python may not be possible.

Scikit-learn integrates well with many other Python libraries, such as [matplotlib](https://en.wikipedia.org/wiki/Matplotlib) and [plotly](https://en.wikipedia.org/wiki/Plotly) for plotting, [numpy](https://en.wikipedia.org/wiki/NumPy) for array vectorization, [pandas](https://en.wikipedia.org/wiki/Pandas_(software)) data frames, [scipy](https://en.wikipedia.org/wiki/SciPy), and many more.

**4**  **RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS**

The result is as the following image below

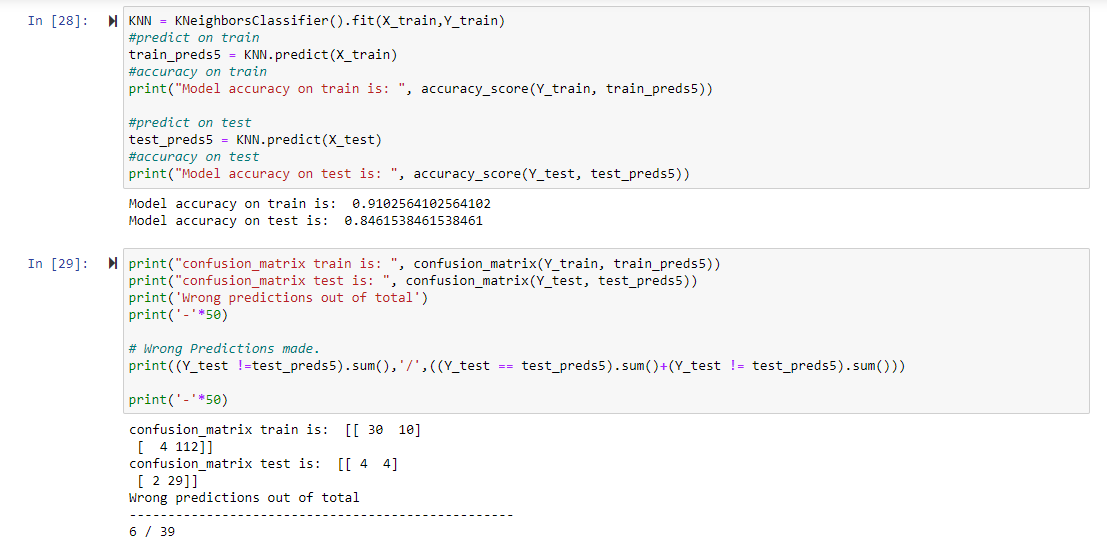
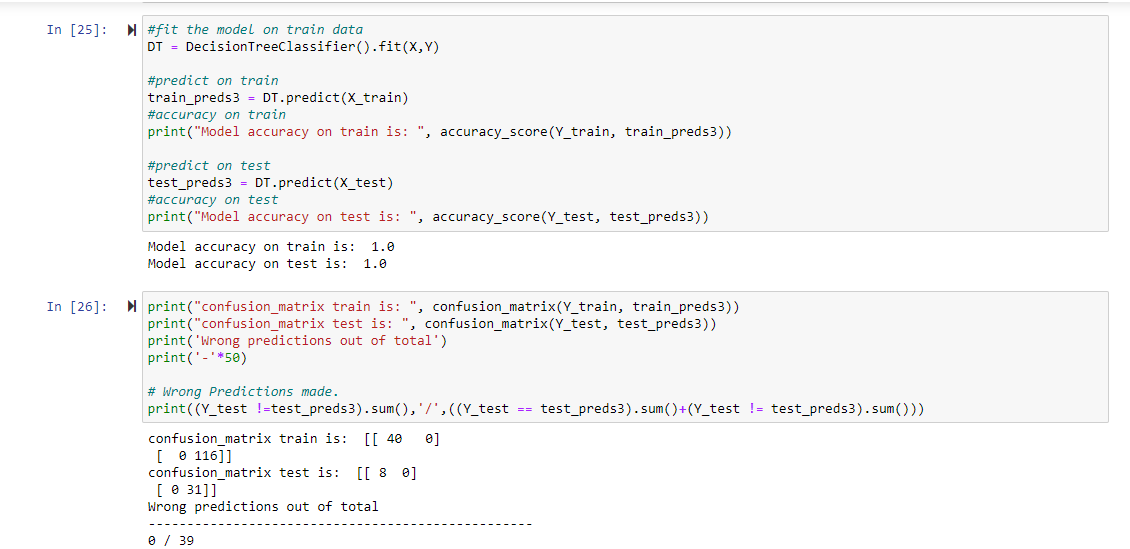
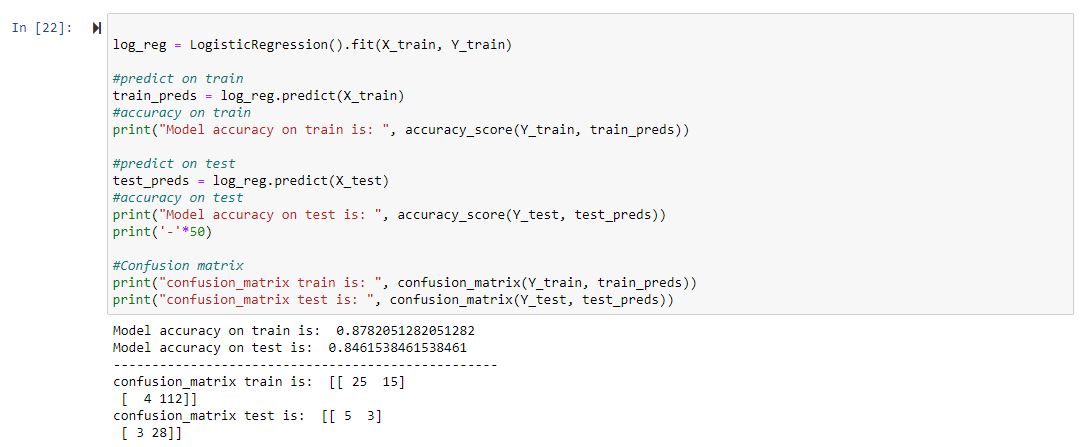
****

Fig 4.1

So by observing the above figure 4.1 we can say that decision tree classifier is performing best out of all three classifier

**5 SUMMARY AND CONCLUSION**

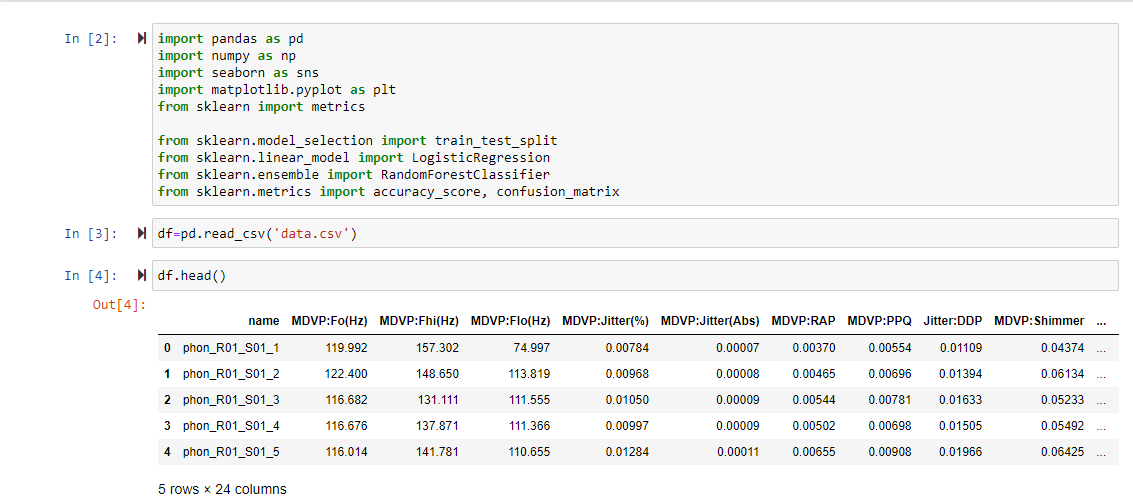
* These robust results suggest that a machine learning approach can indeed be implemented to significantly improve diagnosis methods of Parkinson’s disease. Given the necessity of early diagnosis for effective treatment, my machine learning models provide a very promising alternative to the current, rather ineffective method of diagnosis.
* Current methods of early diagnosis are only 53% accurate, while my machine learning model produces 98% accuracy. This 45% increase is critical because an accurate, early diagnosis is needed to effectively treat the disease.
* Typically, by the time the disease is diagnosed, 60% of nigrostriatal neurons have degenerated, and 80% of striatal dopamine have been depleted.
* With an earlier diagnosis, much of this degradation could have been slowed or treated.
* My results are very significant because Parkinson’s affects over 10 million people worldwide who could benefit greatly from an early, accurate diagnosis.
* Not only is my machine learning approach more accurate in terms of diagnostic accuracy, it is also more scalable, less expensive, and therefore more accessible to people who might not have access to established medical facilities and professionals.
* The diagnosis is also much simpler, requiring only a 10-15 second voice recording and producing an immediate diagnosis.

**REFERENCES**

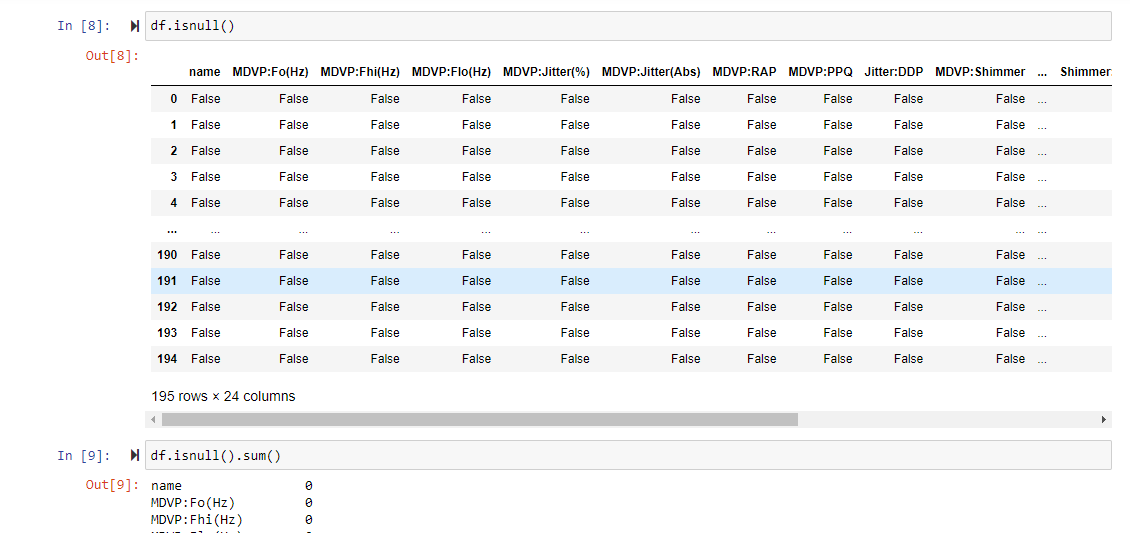
* Bind, Shubham. "A Survey of Machine Learning Based Approaches for Parkinson Disease Prediction." International Journal of Computer Science and Information Technologies 6 (2015): n. pag. International Journal of Computer Science and Information Technologies. 2015. Web. 8 Mar. 2017.
* Brooks, Megan. "Diagnosing Parkinson's Disease Still Challenging." Medscape Medical News. National Institute of Neurological Disorders, 31 July 2014. Web. 20 Mar. 2017.
* Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering OnLine 2007, 6:23 (26 June 2007)
* Hashmi, Sumaiya F. "A Machine Learning Approach to Diagnosis of Parkinson’s Disease."Claremont Colleges Scholarship. Claremont College, 2013. Web. 10 Mar. 2017.
* Karplus, Abraham. "Machine Learning Algorithms for Cancer Diagnosis." Machine Learning Algorithms for Cancer Diagnosis (n.d.): n. pag. Mar. 2012. Web. 20 Mar. 2017.
* Little, Max. "Parkinsons Data Set." UCI Machine Learning Repository. University of Oxford, 26 June 2008. Web. 20 Feb. 2017.
* Ozcift, Akin, and Arif Gulten. "Classifier Ensemble Construction with Rotation Forest to Improve Medical Diagnosis Performance of Machine Learning Algorithms." Computer Methods and Programs in Biomedicine 104.3 (2011): 443-51. Semantic Scholar. 2011. Web. 15 Mar. 2017.
* "Parkinson’s Disease Dementia." UCI MIND. N.p., 19 Oct. 2015. Web. 17 Feb. 2017.
* Salvatore, C., A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M.c. Gilardi, and A. Quattrone. "Machine Learning on Brain MRI Data for Differential Diagnosis of Parkinson's Disease and Progressive Supranuclear Palsy."Journal of Neuroscience Methods 222 (2014): 230-37. 2014. Web. 18 Mar. 2017.
* Shahbakhi, Mohammad, Danial Taheri Far, and Ehsan Tahami. "Speech Analysis for Diagnosis of Parkinson’s Disease Using Genetic Algorithm and Support Vector Machine."Journal of Biomedical Science and Engineering 07.04 (2014): 147-56. Scientific Research. July 2014. Web. 2 Mar. 2017.
* "Speech and Communication." Speech and Communication. Parkinson's Disease Foundation, n.d. Web. 22 Mar. 2017.
* Sriram, Tarigoppula V. S., M. Venkateswara Rao, G. V. Satya Narayana, and D. S. V. G. K. Kaladhar. "Diagnosis of Parkinson Disease Using Machine Learning and Data Mining Systems from Voice Dataset." SpringerLink. Springer, Cham, 01 Jan. 1970. Web. 17 Mar. 2017.

**APPENDIX**

A.Screenshots

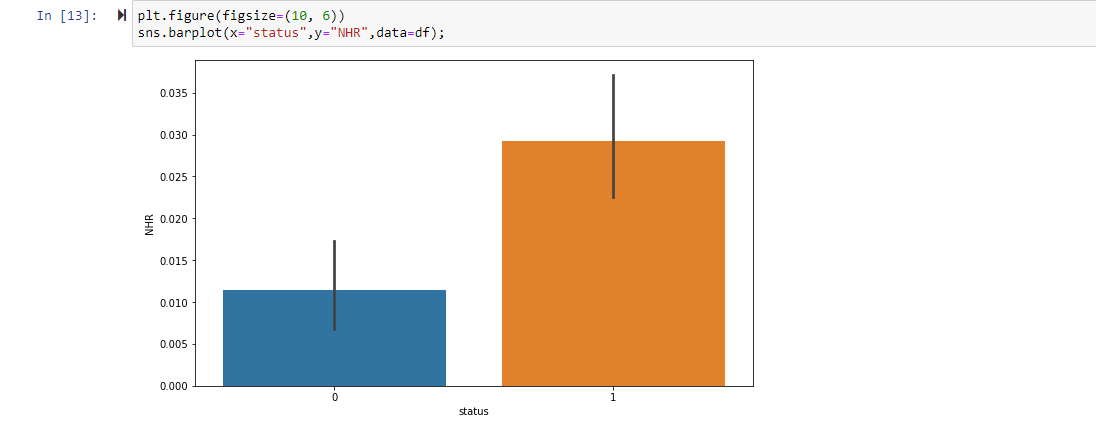


**Fig 5.1** **Importing all necessary libraries**

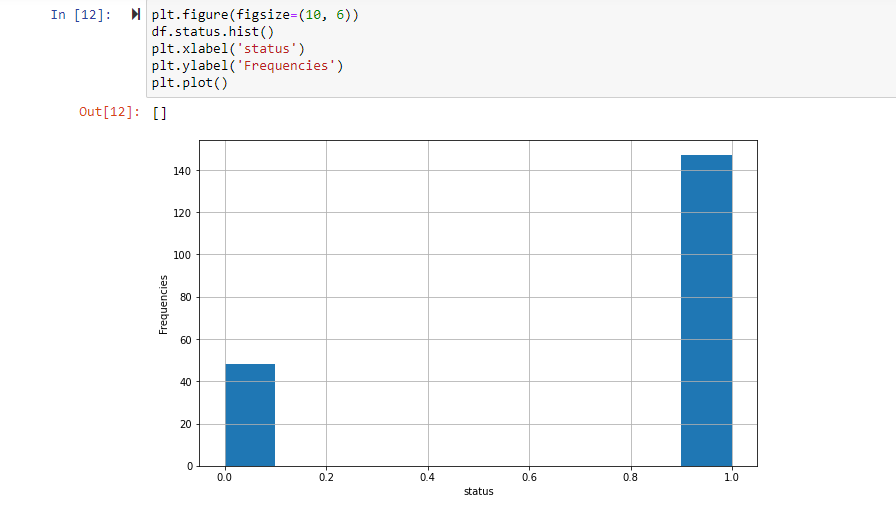


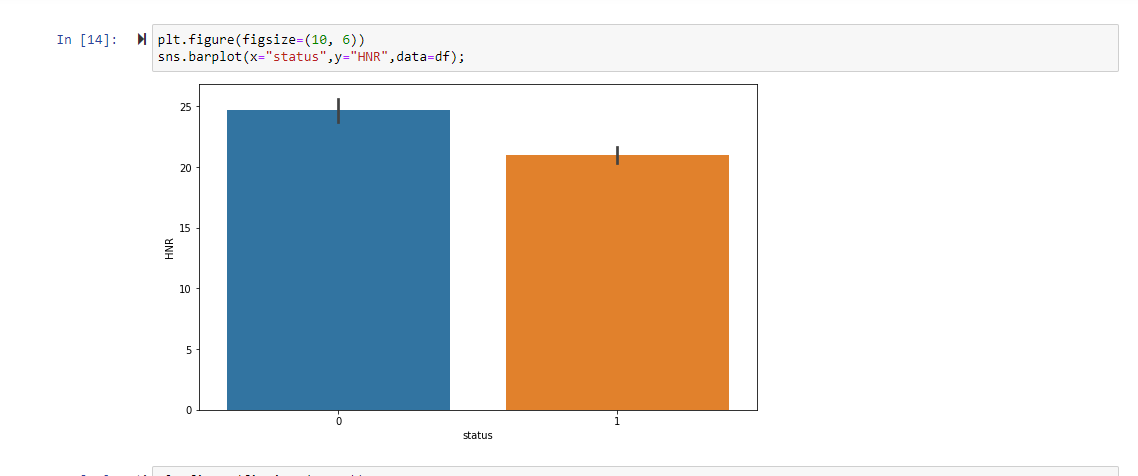
**Fig 5.2 Handling all null values**

**Data Analysis:**

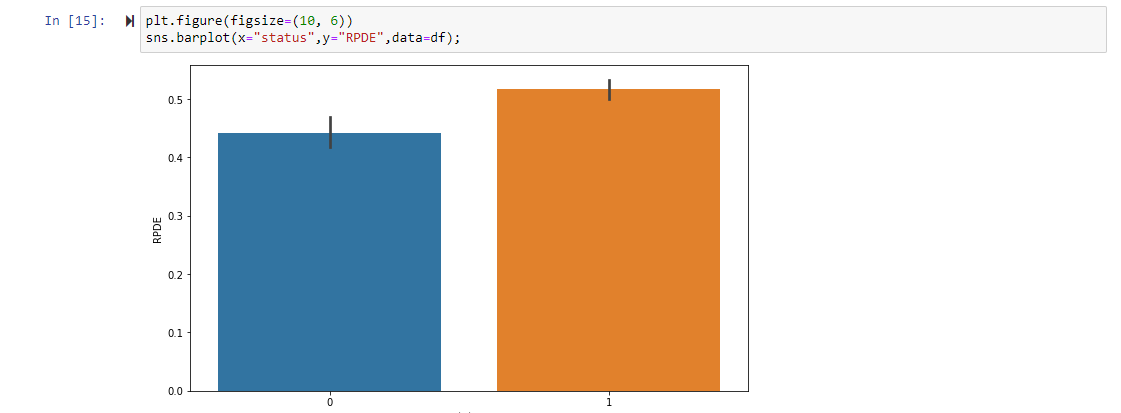


**Fig 5.3 Plotting barplot for status and NHR**



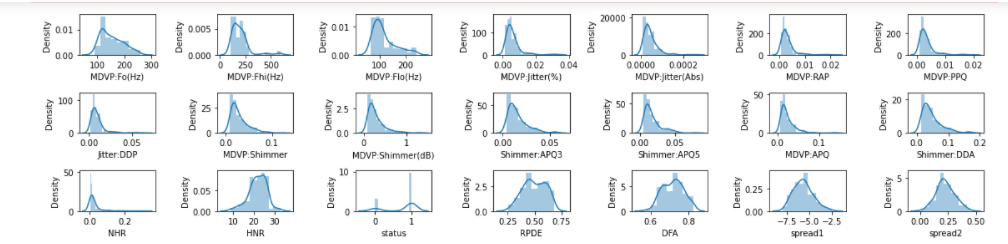
**Fig 5.4 Plotting barplot for status and frequencies** 

**Fig 5.5 Plotting barplot for status and HNR**



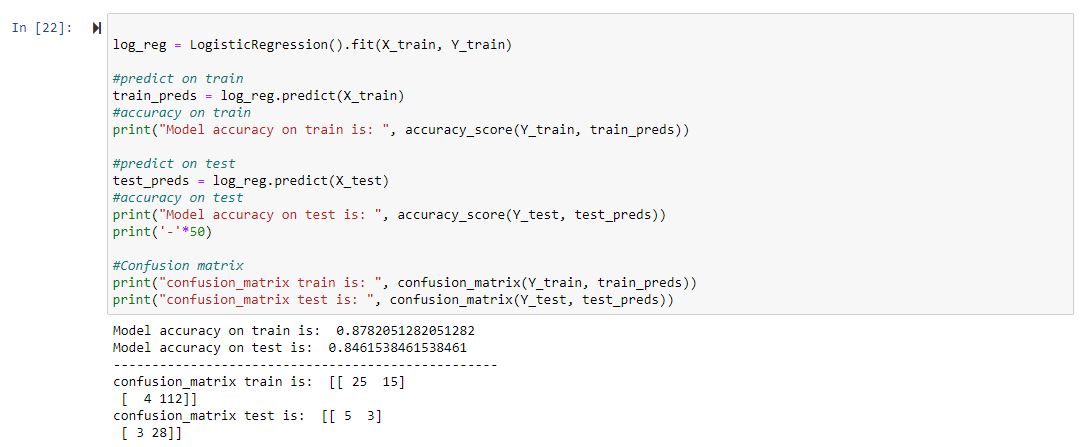
**Fig 5.6 Plotting barplot for status and RPDE**

**Plotting using pair plot**



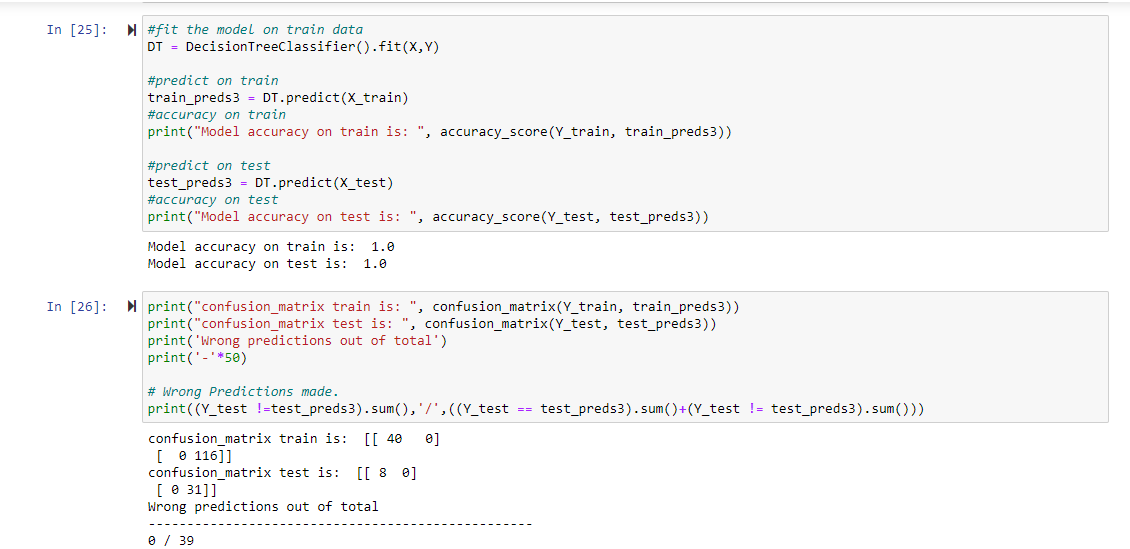
**Fig 5.7**

**Logistic Regression:**



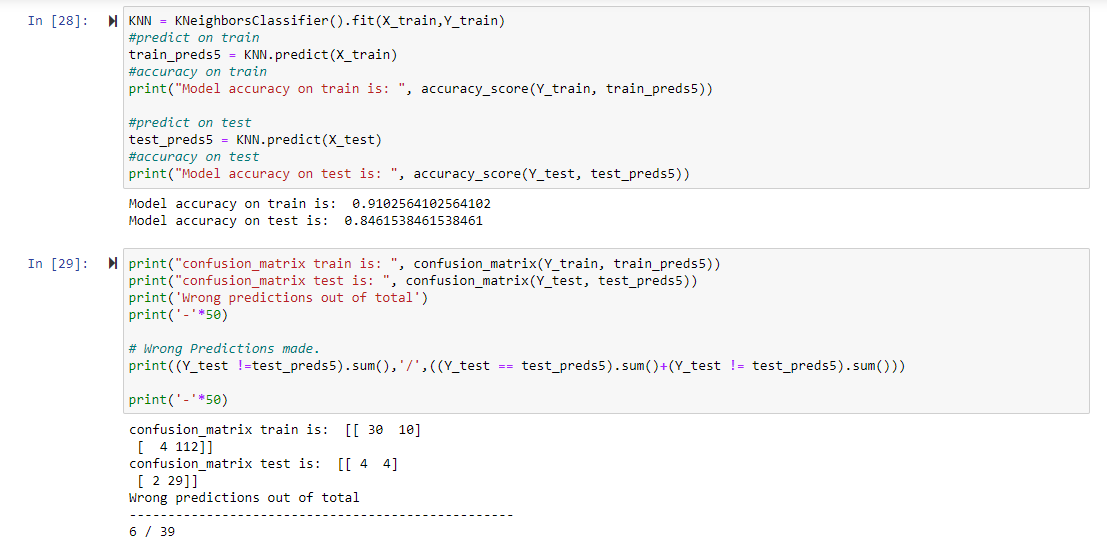
**Fig 5.8**

**Decision Tree:**



**Fig 5.9**

**K Nearest Neighbour:**



**Fig 5.10**

**B SOURCE CODE**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn import metrics

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, confusion\_matrix

df=pd.read\_csv('data.csv')

df.head()

df.shape

df.info()

df.describe()

df.isnull()

df.isnull().sum()

df.columns

df['status']

plt.figure(figsize=(10, 6))

df.status.hist()

plt.xlabel('status')

plt.ylabel('Frequencies')

plt.plot()

plt.figure(figsize=(10, 6))

sns.barplot(x="status",y="NHR",data=df);

plt.figure(figsize=(10, 6))

sns.barplot(x="status",y="HNR",data=df);

plt.figure(figsize=(10, 6))

sns.barplot(x="status",y="RPDE",data=df);

rows=3

cols=7

fig, ax=plt.subplots(nrows=rows,ncols=cols,figsize=(16,4))

col=df.columns

index=1

for i in range(rows):

for j in range(cols):

sns.distplot(df[col[index]],ax=ax[i][j])

index=index+1

plt.tight\_layout()

df.drop(['name'],axis=1,inplace=True)

X=df.drop(labels=['status'],axis=1)

Y=df['status']

X.head()

Y.head()

X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X,Y,test\_size=0.2,random\_state=40)

print(X\_train.shape,X\_test.shape,Y\_train.shape,Y\_test.shape)

log\_reg = LogisticRegression().fit(X\_train, Y\_train)

#predict on train

train\_preds = log\_reg.predict(X\_train)

#accuracy on train

print("Model accuracy on train is: ", accuracy\_score(Y\_train, train\_preds))

#predict on test

test\_preds = log\_reg.predict(X\_test)

#accuracy on test

print("Model accuracy on test is: ", accuracy\_score(Y\_test, test\_preds))

print('-'\*50)

#Confusion matrix

print("confusion\_matrix train is: ", confusion\_matrix(Y\_train, train\_preds))

print("confusion\_matrix test is: ", confusion\_matrix(Y\_test, test\_preds))

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

#fit the model on train data

DT = DecisionTreeClassifier().fit(X,Y)

#predict on train

train\_preds3 = DT.predict(X\_train)

#accuracy on train

print("Model accuracy on train is: ", accuracy\_score(Y\_train, train\_preds3))

#predict on test

test\_preds3 = DT.predict(X\_test)

#accuracy on test

print("Model accuracy on test is: ", accuracy\_score(Y\_test, test\_preds3))

print("confusion\_matrix train is: ", confusion\_matrix(Y\_train, train\_preds3))

print("confusion\_matrix test is: ", confusion\_matrix(Y\_test, test\_preds3))

print('Wrong predictions out of total')

print('-'\*50)

# Wrong Predictions made.

print((Y\_test !=test\_preds3).sum(),'/',((Y\_test == test\_preds3).sum()+(Y\_test != test\_preds3).sum()))

print('KappaScore is: ', metrics.cohen\_kappa\_score(Y\_test,test\_preds3))

KNN = KNeighborsClassifier().fit(X\_train,Y\_train)

#predict on train

train\_preds5 = KNN.predict(X\_train)

#accuracy on train

print("Model accuracy on train is: ", accuracy\_score(Y\_train, train\_preds5))

#predict on test

test\_preds5 = KNN.predict(X\_test)

#accuracy on test

print("Model accuracy on test is: ", accuracy\_score(Y\_test, test\_preds5))

print("confusion\_matrix train is: ", confusion\_matrix(Y\_train, train\_preds5))

print("confusion\_matrix test is: ", confusion\_matrix(Y\_test, test\_preds5))

print('Wrong predictions out of total')

print('-'\*50)

# Wrong Predictions made.

print((Y\_test !=test\_preds5).sum(),'/',((Y\_test == test\_preds5).sum()+(Y\_test != test\_preds5).sum()))

print('-'\*50)