A Project Report on

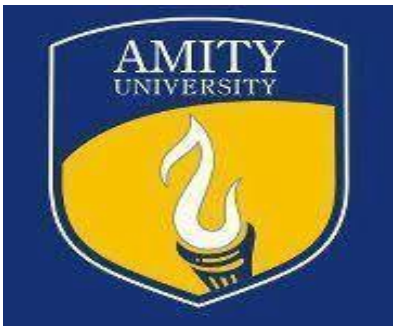
Application of Machine Learning in Recruitment

Submitted To

Amity School Of Applied Sciences

Amity University

Lucknow Campus, Uttar Pradesh



SUBMITTED BY

Aditya Srivastava

M.Stat (Master of Statistics)

A8979220002

4th Semester

External Guide

Mr. Abhyuday Mishra

Abhyutthaan Business Ventures

Lucknow

Internal Guide

Dr. Shailja Pandey

Assistant Professor

Amity University Lucknow

DECLARATION

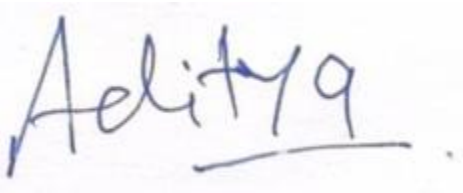
I, Aditya Srivastava student of M.Stat hereby declare that the Summer Internship titled **"Application of Machine Learning in Recruitment"** Which is submitted by me to Department of Statistics, Amity Institute of Applied Sciences, Amity University, Uttar Pradesh, Lucknow, in partial fulfilment of requirement for the award of the degree of Masters in Statistics, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Place: Amity University, Uttar Pradesh Lucknow

Date:

Name and Signature of the Candidate

Aditya Srivastava



CONTENTS

CERTIFICATE

On the basis of declaration submitted by Aditya Srivastava student of M.Stat , I hereby certify that the Summer Internship titled **"Application of Machine Learning in Recruitment"** which is submitted to Department of Statistics, Amity Institute of Applied Sciences, Amity University, Uttar Pradesh, Lucknow, in partial fulfilment of requirement for the award of the degree of M.Stat, is an original contribution with existing knowledge and faithful record of work carried out by him under my guidance and supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Date: 30.05.2022

Dr. Shailja Pandey
Assistant Professor
Department of Statistics
Amity Institute of Applied Sciences
Amity University, Uttar Pradesh, Lucknow

Dr. Asita Kulshreshtha
Head of Institute
Amity School of Applied Sciences

CERTIFICATE

This is to certify that Mr. Aditya Srivastava had been engaged with the company M/s Abhyutthaan Business Ventures Pvt. Ltd. from 03-01-22 to 15-04-2022. During his tenure, he worked on Application of Machine Learning in Recruitment

We wish him luck for his future endeavors

Issued by

Authorized Signatory

For

Abhyutthaan Business Ventures Pvt. Ltd.

Date: 15/04/22

Place: Lucknow, Uttar Pradesh

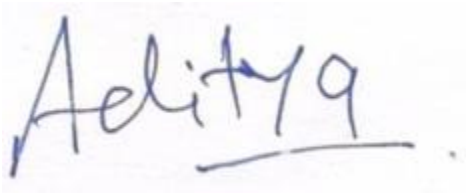
ABHYUTTHAAN BUSINESS VENTURES PVT LTD
7A-1A DALIBAGH, TILAK MARG
LUCKNOW
UTTAR PRADESH

+91 522 4101800 | +91 84000 5 8400
business@abhutthaan.com
www.abhyutthaan.com

PLAGIARISM DECLARATION

I, Aditya Srivastava, student of Master of Statistics, hereby declare that the contents of this project report titled **"Application of Machine Learning in Recruitment" is** a genuine work of mine, submitted as Major Project/Dissertation, under the Post Graduate program of Master of Statistics. And this internship project has been evaluated by online anti-plagiarism software. The contents and material were found satisfactory and within the permissible limit of content copied.

Name and Signature of the Candidate



PLAGIARISM REPORT

DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

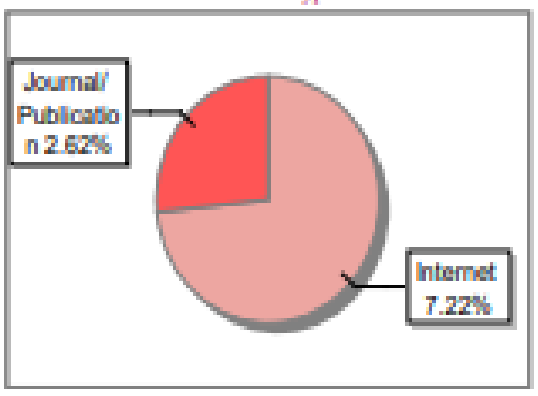
Submission Information

Author Name	Aditya
Title	Application of machine learning in recruitment
Paper/Submission ID	539591
Submission Date	2022-06-09 14:20:19
Total Pages	38
Document type	Project Work

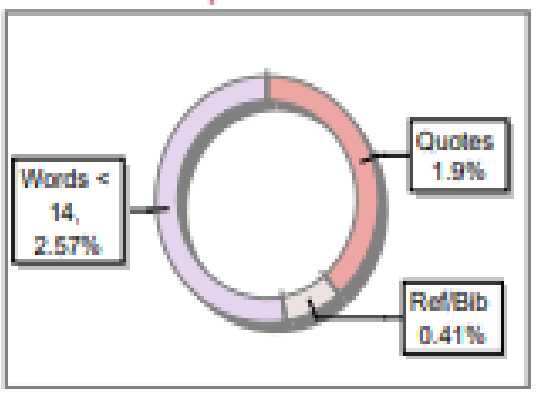
Result Information

Similarity **11 %**

Sources Type



Report Content



Alert..! Improper usage of 'Quotes' work in a report

Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Sources: Less than 14 Words Similarity	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

ACKNOWLEDGEMENT

I express my deep sense of regard and gratitude to my teacher and guide, Dr. Shailja Pandey, Amity school of applied sciences, under whose supervision I had been conducting this work. I will always be grateful to her for his continuous encouragement, valuable suggestions, and generous support. I wish to acknowledge my sincere thanks to all respected teachers, especially Dr. Gunjan Singh, department of Statistics, Amity School of Applied Sciences, Amity University for her cooperation and guidelines throughout this course. All the members of the library, and data center, my sincere thanks.

Finally, I take this opportunity to thank Almighty God and my Parents for providing me with the opportunity to be a part of Amity University.

1. Abstract

Recruitment refers to the overall process of identifying, attracting, and interviewing suitable job candidates. It is a process involved in choosing an individual for paid and unpaid roles. Nowadays recruitment can be done through online and offline modes and due to covid-19 outbreak online mode of recruitment is very efficient and helpful to do and our paper is based on recruitment based on online modes. In this project, we discuss the application of machine learning in recruitment. Machine learning, with to the given tasks with some predefined algorithm. We cannot handle it manually such a large amount of data therefore we need to do it through machines. We are to going how online recruitment can transform the ways of recruitment. Some factors will affect the recruitment of the candidate or what does a company need from the candidate.

2. Introduction

Earning money is the necessity for earning to fulfill our daily needs that a person must do the job whether it is in the government sector or the private sector. For doing the job we have to clear the recruitment process. Recruitment is the way through which any company is selecting a candidate for a particular post. It is the way by which any candidate gets a job opportunity. It is a multi-process system. Shortlisting a candidate can be an arduous process. Recruitment can be done through two modes online and offline modes. We are focusing on the online recruitment system. The online recruitment system is immensely helpful to do the recruitment in a limited amount of time. They help in the process in such a way that each step takes so little time that it will help a company to save that was wasted when the recruitment is done through offline modes. Online recruitment system helps recruiters and candidates both in many ways. It can help to provide an honest candidate for the recruiter and a good job for the candidate. In this project, we are showing how online recruitment is done through machine learning and how machine learning is transforming the whole recruitment system.

Machine learning is a branch of artificial intelligence that helps to work through algorithms that will make upgrades automatically through experience using data. Machine learning can assist the recruitment process in many ways. In talent sourcing, CV screening, job advertising, candidates' preselection and assessment, candidate engagement, and predicting hired needs. For recruitment, data is collected from many sites or in many ways such as job portal sites which are Naukri.com or InternShala or LinkedIn, etc. Through machine learning candidate is selected for CV screening which has the same qualifications required for the post. Companies collect the data for CV screening from various social media sites like Facebook etc. These sites help to enhance the recruitment process.

3. Dataset

It is the collection of data. Everything around is the data. What we use on mobile has a collection of data. What we do in a day is also data. So, if we notice everything has data. Data is the most important thing to do in any analysis. If we do not have the data then we cannot make any predicted conclusions with any data. Prediction with data helps to understand the outcomes of the particular objective.

In this project, we have the data of recruitment which is online recruitment for the post of assistant manager. data has 300 candidates for the post of assistant manager educational qualifications are required to hire any candidate. the data has gender knowledge, their high school marks and boards, intermediate marks, and boards, and from which stream they have passed the exam. Their graduation details and their post-graduation details. Apart from educational qualifications if they have work experience in months and they have passed a written exam for the post the assistant manager their marks were also in the data. The whole data was collected by the company which is Abhyutthaan Business ventures and then the analysis is done on the dataset.

4. Recruitment

Recruitment is a method to determine the number of vacancies, analyze the job requirements, review the participants, screen, and then select the perfect candidate for the job. The company must see that the candidate is completely eligible for that job and by his qualifications, knowledge, extra activities, hard-working, consistency, compliance, punctuality, technician's calculation, and theoretical knowledge. According to **EDWIN B FLIPPO**

“RECRUITMENT IS THE PROCESS OF SEARCHING THE CANDIDATES FOR EMPLOYMENT AND STIMULATING THEM TO APPLY FOR JOBS IN AN ORGANIZATION”

Recruitment can be done in two ways: -

1) **Online Recruitment System** – online recruitment system is the method of recruitment system which is done through an online portal such as job portal sites or social media sites which is easy. It saves a lot of time, money, and the services of the company and recruiters.
2) **Offline Recruitment System** – it is the traditional system which was doing recruitment for ages. It takes a lot of time, money, and services. Their medium of advertisement was posters, newspapers, s, etc. nowadays it is replaced ban y online recruitment system. Hardly there is any company left which are using the method for recruitment because not even the private sector but also the government sector is also using online recruitment as it maintains security.

5. Online Recruitment System

Due to the pandemic and adapting to the digital world, everything is now going in online mode. People prefer to work from the home mode which enhances the online recruitment system. In an online recruitment system recruiter have to post a job or vacancy on social media sites or job portal sites where the searching candidates can submit their Resume or CV’s so that the company will their profile. Some job portal sites are LinkedIn, Internshala, Glass, etc. and many other sites help a person to find a suitable job for them. In the job portal sites when a person finds a suitable then the candidates submit the CV’s and the candidate's work is done and then a recruiter will see the candidate’s CV and all the information given in their profile on ton sites and after seeing the all the things it is decided whether the candidates are selected or rejected for the further round which is the assessment round. All the processes in this system are done online as it will help the recruiter and candidate to work efficiently.

5.1 Advantages of an Online Recruitment System

1) The candidate and the company don't need to belong to the same city. So online recruitment system helps candidates to connect with the company even if their distance is long.

2) Online recruitment system saves a lot of time which spends in the starting process of recruitment if it is done in the offline mode. After the candidate submits their CVs, a recruiter can see the profile and if it is a deserving candidate then the person is selected for the further round which saves a lot of time as it is very time-consuming to do for all the candidates at once.

3) The online recruitment system is a very cost-effective process. When the candidate is hired the company will give them training for some months and also during the recruitment process there are many ways in which money can be spent if it is not done in online mode. During the onboarding, training, and retention a lot of money is saved by the online recruitment system. If the company’s turnover rate is high within a short period, then the company has to hire again which then again also spends a lot of money.

4) Since the process is online so there is a large number of candidates gathered around and it is also not important that they belong from the same country. Nowadays social media platforms and job portals are very popular and there is a huge audience looking they're for the job, so it helps and allows and company to find the right candidate for the job.

5) Any recruiter can make an advertisement for the vacancy and they wanted to be user-friendly so that no one has to face a problem understanding it. A recruiter has to make sure that the form should not be large otherwise candidates get tired and do not want to fill that form and wanted to skip that. For the applicant, the process should also be easy and clear just to submit their CVs and fill a small form with limited questions so that it also be user-friendly.

6) In an online recruitment system, the recruiter makes a job ad more dynamic. They want to be more creative with their advertisement. They will attach a video with the ad and also explains the company which helps any applicant to know about it in detail.

7) In an online recruitment system, the internet will give us a flexible environment which helps the recruiter otherwise if the vacancy of the job is given in the newspaper, then the company has to pay for every post, and once, they given the ad they cannot amend or edit that advertisement. The advertisement will print the way they have submitted and suppose they hire someone they do not want another ad in the newspaper but they have paid the fees or the wrong information is given in the newspaper then it cannot be edited. But in the online system, we can edit, amend, or if we want to delete then we can simply do we do not have to pay for that.

8) The online recruitment system helps us to shorten the hiring process by shortlisting the applicant after they submit their CV's so that they will not have to take an assessment, not for many applicants and this will help also in the further round.

9) If any organization wants to keep their things privately, they will prefer an online recruitment system. When the job vacancy or the organization is very sensitive or controversial, they prefer to be online. These employers can conceal the name of their firms or even search and contact candidates directly without the need of posting a job advertisement.

10) The online recruitment system helps us to improve hiring efficiency as the recruiter and applicants want to do the process in a user-friendly manner.

5.2 Process of online recruitment system

For any company, the recruitment process is a multi-step process which is shown in the given below: -

1) It is the first in which the recruiter identifies job requirements in any department of the company.

2) Now the second step is to prepare for the job description and job specification and all the important things that they should know about the vacancy.

3) Now the third step is to advertise that job on any social networking sites, or job portal sites so that they can attract as many as candidates. The advertising part is very important as the company did not want that a good candidate missed seeing the job vacancy and will be a loss for a company to lose a valuable asset.

4) Now the fourth step is to attract as many as candidates so that more candidates will fill the form and the company will have more options to select.

5) In the fifth step when the candidates submit their CVs or Resume then the recruiter has to shortlist some candidates to move further.

6) In the next step, the recruiter analyzes the candidates in academics and other fields which are related to the job and tests the candidates in all sectors of the field.

7) the last step is to shortlist the candidate for the job by taking interviews which are taken by the HR department and the technical round and then the most perfect candidates will be selected.

The whole recruitment process is not a one-day or two-day process but it is a long process which is taken place for at least a month or more than that. The assessments and the interviews are very important for any candidate as many candidates will pass the document checking but they will get rejected if they do not have sufficient knowledge of their subject. In the interview part, the panel can be of one person or it can be two, three, or even five persons which are inspecting the candidates in different sectors.

5.3 Problems facing during online recruitment system

There are also problems related to the online recruitment system which are as follows: -

1) Depending upon the job portal sites the company has to pay a subscription fee to post a job or to track the applicants. Sometimes we invest more but we do not get the results that we wanted and this will be a loss

when it takes time to get a good applicant for the role. The costs will increase which is not good for any company.

2) During the interview, if the recruiter has a bad network connection or there is an electrical fault or any technical fault then it will be an embarrassing moment for the recruiter. This will somehow damage the reputation of the company.

3) Sometimes during the online recruitment process communication will between the recruiter and the candidate becomes a problem. It can lead to misconstrued or wrong messages to the recruiter and the applicant can be rejected for that. However, in offline recruitment applicants and recruiter are face to face so they cannot have communication whatever the applicant says the recruiter will understand perfectly without any communication fault.

4) In the online recruitment system when applicants fill out the form and submit their CVs there will be some applicants whose applications might be fake and who wanted to get information about the company or hiring officer.

5) When the company receives a large number of applicants or job advertisements need to be constantly then it will create a disturbance for the employer who monitors all the things and then the employer's work is also disturbed so it needs someone to supervise closely.

6) Not all apps or services offer an in-depth analysis of the advertisement therefore it will become difficult to understand what is working or what's not and how to optimize our ad.

7) When the candidates see the advertisement then some of the candidates will fill the form without, they see about the requirements and thus not all candidates are good some bad candidates too will come, and it is the employer's work to find out the good candidates.

8) Since there are a lot of applicants applying for the job so it will chaos if it will not manage properly but for the candidates, it will result in a lot of competition and which to get a job harder.

6. Machine learning

Machine learning is a very demanding technology nowadays. It is a subset of Artificial Intelligence. It was first named by **ARTHUR SAMUEL** in 1959 who is an expert in computer gaming and artificial intelligence stated that **"IT GIVES COMPUTERS THE ABILITY TO LEARN WITHOUT BEING EXPLICITLY PROGRAMMED"** but the first definition was stated by **TOM M. MITCHELL** which said that

"A COMPUTER PROGRAM IS SAID TO LEARN FROM EXPERIENCE E CONCERNING SOME CLASS OF TASKS T AND PERFORMANCE MEASURE P IF ITS PERFORMANCE AT TASKS INT, AS MEASURED BY P IMPROVES WITH EXPERIENCE".

In our real-world examples such as score prediction, spam detection, and gaming we all use them in our daily life. We as a human is very independent of machine learning methods. From morning to night, we use these methods but do not know that we are using them. We are so much dependent on this system that we do not know how to do that kind of work if we do not have that type of help. Machine learning helps mankind for a better future but with so much help around even if we do not want that our nature and we cannot think without their help. Everything has merits and demerits so does the machine learning which we discuss later.

Machine learning is within computer science, but it is different from traditional computational approaches. In traditional computational approaches, algorithms are explicitly programmed instructions given to the computers whereas in machine learning computers are allowed to train on data inputs and use statistical analysis in building the model for the sample data.

In machine learning when the input the data it will automatically interpret the data to analyze the data using some machine learning algorithms.

According to Wikipedia, in the next ten years machine learning can generate up to one billion dollars for the pharmaceutical company and it is estimated that it can replace 25% of the jobs which is a great disadvantage.

6.1 Need for machine learning

In the present scenario, every other person is using the data. Therefore, it creates a lot of amounts of data. The requirement for machine learning is increasing day by day. As a human, we always have a limitation. We cannot deal with large or big data manually it would not be possible to work manually. So, to overcome the problem of handling data we need proper algorithms and computer systems for this machine learning make will make things easy.

We can teach machine learning algorithms by giving them a huge amount of data for traversing the data, building the model, and predicting the result from the model. The efficiency of the model can be determined by the cost function. Through machine learning, we can save both time and money.

There are important factors that show the importance of machine learning –

1) The rapid increase in the production of data
2) Solving complex problems that are difficult for human
3) Decision-making in various sectors including finance.
4) Finding hidden patterns and extracting useful information from data.

According to some research, we can create around 2.5 quintillion bytes of data in a single day and it also says that by 2020 we can create 1.7 megabytes of data every second which can result in a lot of data. With so much data we can finally use it to analyze and make predictive models and in real-world examples.

"**Just as electricity transformed almost everything 100 years ago, today I have a hard time thinking of an industry that I don't think AI will transform in the next several years**" by **ANDREW NG**.

6.2 History of Machine Learning

a) The Early History of Machine Learning (pre – 1940)

In **1834**, Charles Babbage the father of the computer come up with a device that was programmed with punch cards. However, the machine was never built, but today all modern computers relied on that logical structure. In **1936**, Alan Tuning had a theory of how machine can decide and execute a set of instructions.

b) The Era of Stored Program Computers

In **1940**, the first manually operated computer, **ENIAC** was invented which was the first general computer. After this **EDSAC** was also made in **1949** and **EDVAC** in **1950** for the stored-program computer. In **1943**, a human neural network was modelled with an electrical circuit. From now on scientists started applying the idea of human neurons by analyzing how they work.

c) Computer Machinery and Intelligence

In **1950** Alan Tuning published a paper **"Computer Machinery and Intelligence"** on the topic of artificial intelligence. In this project, he asked the audience how can machines work.

d) Machine Intelligence in Games

In **1952**, A program was created that helped an IBM computer to play a checkers game. It was created by Arthur Samuel, who was the pioneer of Machine Learning. It performed better more it played. The term "Machine Learning" was first spelled by Arthur Samuel in 1959. The First "AI" Winter. The duration **1974** to **1980** is called "AI" Winter because it was the tough time for AI and ML researchers.

"In this duration, failure of machine translation occurred, and people had reduced their interest in AI, which led to reduced funding by the government to the research."

e) Machine Learning from Theory to Reality

To remove echoes over phone lines the first neural network was applied to a real-world problem using an adaptive filter in **1959**. A neural network NET talk was invented in **1985**, which was able to teach itself how to correctly pronounce **20,000 words in one week**. It was invented by **Terry Sejnowski** and **Charles Rosenberg**. In **1997** IBM invented the intelligent computer named deep blue, which won the chess game against the chess expert **Garry Kasparov**. It became the first computer that had beaten a human chess expert.

f) Machine Learning in the 21st century

In the year **2006,** a new name has given to neural net research as **"Deep learning"** by computer scientist **Geoffrey Hinton**. Nowadays, it has become one of the most trending technologies. A deep neural network was created by Google in **2012** which learned to recognize the image of humans and cats in YouTube videos. In **2014**, Chabot &

quot; Eugen Goostman & quot; cleared the Turing Test. It was the first Chabot who convinced 33% of human judges that it was not a machine. In **2014** a new deep neural network DeepFace was created by Facebook claimed that it could recognize a person.

g) Machine Learning at Present

Machine learning has not a new advancement in research. It is present everywhere in engineering, medical fields, business, technology, etc. such as self-driven cars, amazon Alexa, transportation, and commuting, etc. it has to mow many ways of algorithms such as supervised machine learning, unsupervised machine learning, support vector machines, reinforcement machine learning, clustering analysis, association, etc.

6.3 Features of machine learning

There are some features of machine learning which are as follows

1) Machine learning uses data to predict various patterns and trends in data.

2) It does not require any human to do every work. It can learn from experience and improve it automatically.
3) It is a data-driven technology. Every work we do or intend to do we cannot do without data. There is no meaning of algorithms or any other things if the data is not present.

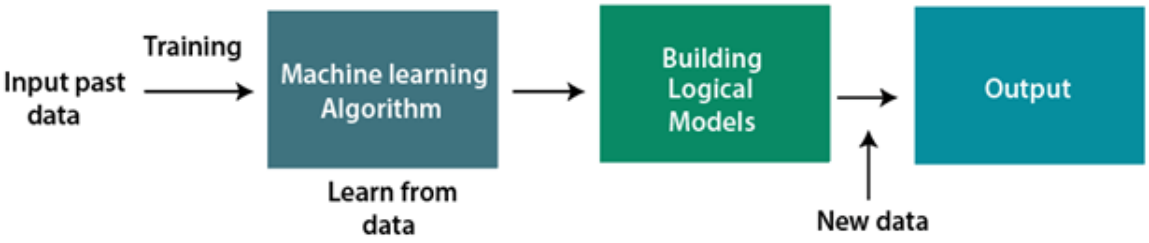
4) Machine learning is common to data mining. Therefore, it deals with a big amount of data.

6.4 How does machine learning work?

Machine learning learns from historical data and builds prediction models, it also modifies the model as the new data comes and predicts the updated outputs. The accuracy of the data depends on the huge amount of data. If the data is big then the model is also better and the prediction is also more accurate.

If we have a complex problem and we have to find some predictions, so rather than writing a whole code, we just need particular algorithms and new data to build an efficient model.

The block diagram of how machine learning works are as follows



in the block diagram shown above, we have to first input the whole data which was collected through the past experiences after the data is split into training data and testing data. In the training, the data model is created and in testing, data prediction is made. Now through the training data model is created different machine learning algorithms and output is then created but if the new data is added then it modifies them and then again it gives the output.

6.5 Applications of Machine Learning

There are some applications in machine learning such as: -

1) Netflix uses machine learning in recommendations. When someone opens Netflix app for the first time it shows infamous recommendations on TV shows and movies but the data is collected when the person is using the app and after then recommendations in their app.

2) A facial recognition system is present in every mobile phone in which through patterns it recognizes the person which is an application of machine learning and natural language processing.

3) Facebook has an auto-tagging feature in which when the person is uploading the photo then through machine learning if the person is a user of Facebook, then it automatically tags that person.

4) Google has a spam filter in Gmail through machine learning and natural language processing it differentiates between spam and non-spam mail.

5) Amazon's famous Alexa is an example of natural language processing and machine learning is an advanced version of virtual assistant that not only plays songs or movies but also talk to the person and book a cab, connects with other IoT devices at home, office, track your health and comes in many different languages.

6) Self-driven cars are also an example of machine learning in which the car itself drives, stops in signal and when some car comes towards it then it will automatically shift from left to right.

7) Google legal translate is an example of machine learning and natural language processing which helps any person to transform their text, image, or written by a person in any language which is not important that their language is either a foreign language or a regional language. It will transform that text or image into whatever language they want.

8) Dynamic pricing is another example of machine learning which helps the company to set the prices of a product automatically. if the demand is high the price is also high but if the demand is less than the price is also less. It is a great help for the company to not set the prices of every item now and then.

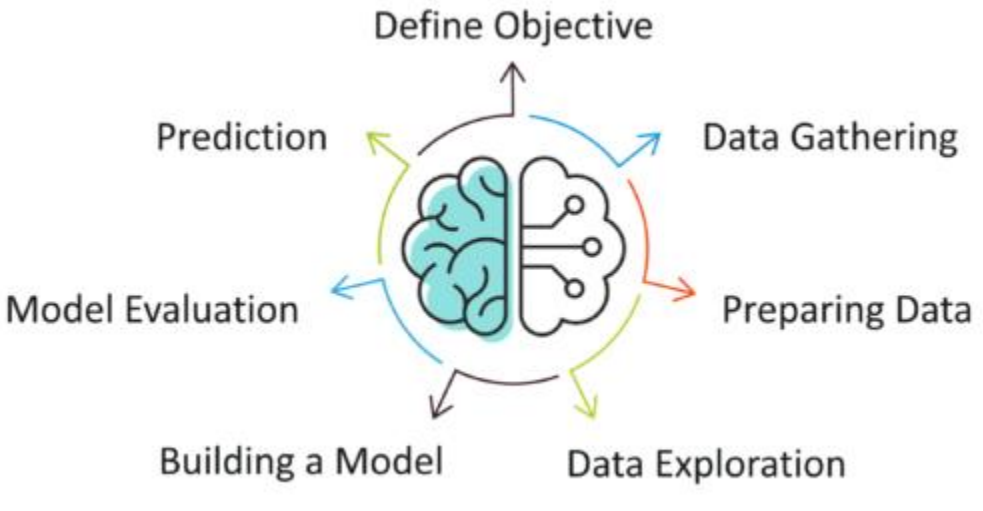
9) Transportation and commuting are also good applications of machine learning. Uber company uses this method for booking cabs or auto. When any person books a cab then within some seconds the location was sent to the cab driver who is nearest to them and within some minutes the cab will arrive. This makes the process of going anywhere in the city a lot easier. The booking person only has to give their arrival and destination address to the application and it will directly link to the nearest cab which can arrive as soon as possible.

10) Google maps are a reliable source nowadays. They use machine learning and natural language processing technique to find any location. they also give traffic alerts which help the person to avoid going on that road

and instead of choosing another road. It will navigate the whole by saying in any languages they want and gives all the directions which the person needed to go.

6.6 Process of Machine Learning

The various process of machine learning is given as follows: -



1) Define objective -The first and the foremost important step is to obtain an objective. The objective should be very clear otherwise the prediction gets wrong. We should complete an understanding of what is going to be predicted.

2) Data gathering -When we think about gathering the data, we should know what are items that we require for prediction and from where? Data can be gathered manually through a questionnaire which can be filled out by many people or through the web where there are dozens of data is present therein that we can work on.

3) Preparing data -When data is found it cannot be in the right form. There are always some missing values or some irregularity showing in the data we should have erased before we start the work on data. Data mustn't have any missing values.

4) Data exploration -Data exploration is the important part. In this stage, we have to fully understand the data with all the patterns and trends of data. All the important findings in the data are found here and the correlation between variables is understood.

5) Building a model -All the importance which can be extracted from data can be found in the data exploration stage. In this stage, we divide the data into training data and testing data. In the training data, we build the model and analyze it.

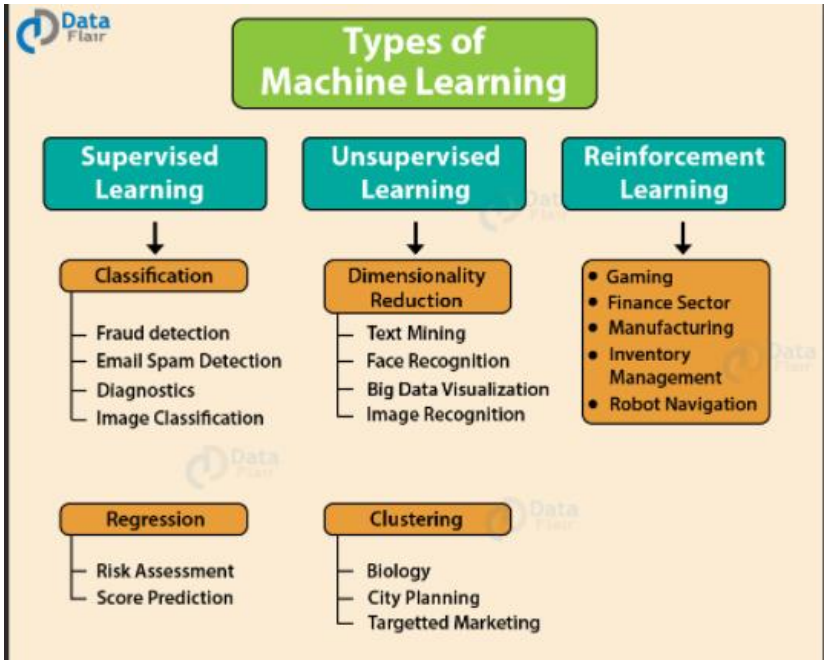
6) Model evaluation -In this stage, we predict the outcome of the testing data. We will check the efficiency of the model and how good a model is. When the accurate model is selected then we proceed further in which we implemented parameter tuning and cross-validation is used in the model to improve it further.

7) Prediction -when the model is evaluated and improved then it is time for a prediction. The final result can come either in a categorical variable that is TRUE or FALSE or in a continuous variable that is the value of a stock.

6.7 Types of machine learning

Machine Learning helps in many ways.

There are three types of machine learning present they are as follows: -

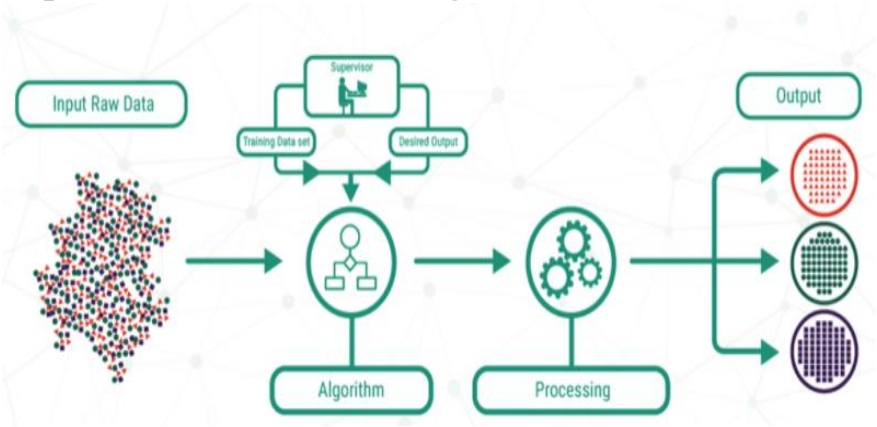


1) Supervised Learning which uses labelled data and gives the labelled output through algorithms that can be done through either classification (for example fraud detection, email spam detection, diagnostics, image classification, etc.) or regression analysis (for example risk assessment, score prediction).

2) Unsupervised Learning uses unlabeled data and gives the output in the cluster form which is the unlabeled output through algorithms either through cluster analysis (for example biology, city planning, targeted marketing) or through dimensionality reduction (for example text mining, face reduction, big data visualization, image recognition).

3) Reinforcement Learning does all the work by itself through algorithms. For example, gaming, finance sector inventory management, manufacturing, and robot navigation.

a) Supervised machine learning

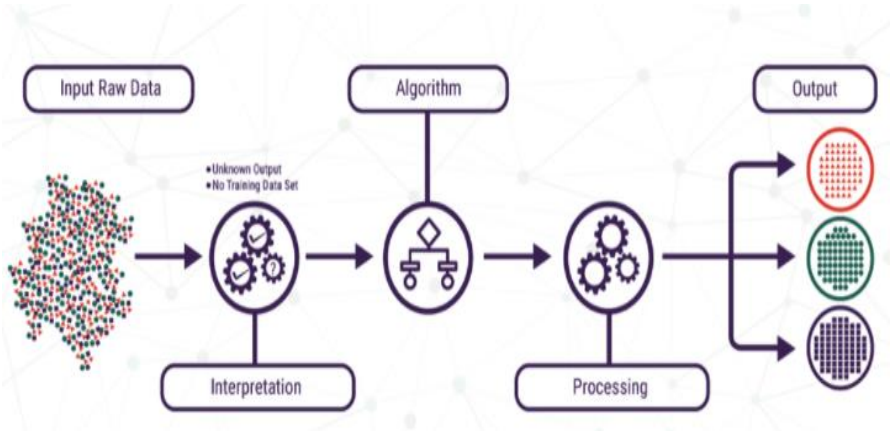


Supervised learning works on the labelled data in which we have input and also the desired outputs. It states that **"SUPERVISED LEARNING IS A TECHNIQUE IN WHICH WE TEACH OR TRAIN THE MACHINE USING DATA WHICH IS WELL LABELED"**

In supervised learning, the machine has the labelled input and desired output. The process is when the labeled data is entered in the machine then it applies algorithms which is the machine learning methods such as regression analysis and classification in this stage data exploration analysis is done then data is split into training and testing data and then it builds the model and evaluates it on the training data and then it predicts from the testing data. After predicting it will compare the output with the desired output and then it will modify the model accordingly.

In this type of machine learning, some guidance system is required to enable the task and since we know the desired output, the output will be some in the form which is almost similar to the desired output.

b) Unsupervised machine learning



In unsupervised learning, it states that

"UNSUPERVISED LEARNING INVOLVES TRAINING BY USING UNLABELED DATA AND ALLOWING THE MODEL TO ACT ON THAT INFORMATION WITHOUT GUIDANCE"

In this type, unlabeled data is used and it does not have the desired output and the output is the unlabeled output it comes in the cluster form. The process of unsupervised learning is that when unlabeled data comes into the

machine it is first interpreted by doing exploratory data analysis and then algorithms which are machine learning methods such as clustering analysis, association analysis, and hidden Markov model which are used to build and evaluate the model by training data and then it predicts the model using testing data.

This type of machine learning machine does not require any guidance which was needed in supervised machine learning and since the machine does not know anything about data therefore it will give an unlabeled output which is in the cluster form.

7. Machine learning in recruitment

Hiring an employee is not an easy task. It takes weeks or even months to get a perfect employee for the company but machine learning this process becomes easy and it will not take weeks or months but it will do it very quickly. in the normal hiring process candidate's screening takes a lot of time as it is done manually. In the screening, process candidate is tested on their qualifications and if it is good then they will be shortlisted which is a very short process if we do it through machine learning but if we do through manually then it will take a whole lot of time.

There are some ways in which machine learning can help in the recruitment process are given as:

1) If recruiters find some problem related to the hiring process it will be solved through machine learning for that they should know the objective of the problem and then they apply machine learning to them to solve it.

2) Machine learning helps in the hiring process to make a strategy for evaluation after screening is done the candidates are shortlisted for the examination. It tracks the evaluation process which is stored by the recruiter in the machine.

3) Machine learning helps in gathering the data and after it will refine, format, combine and process the data in the possible way which saves a lot of time.

4) Machine learning is done through machines which means that the precision level is very high as in the normal hiring process human error is normal and it will result in a low precision level which can change the outcome of the result.

5) Recruiters can make what type of model they want and they set up the algorithm's as they want in want to achieve better results.

6) Machine learning helps the company in the cost reduction in the hiring process rather spent a lot of money in the hiring process if it is done manually.

8. How machine learning is revolutionizing recruitment?

Many challenges come across during recruitment but one of the main challenges or say problems is the candidate's numbers. There are a massive number of applicants coming for the job but the recruiters cannot hire all of them

and to do the process manually it takes a lot of time, resources, and effort which affects other aspects too so it is better if the process is done through machine learning.

1) Machine learning helps to get applicants' information -Lots of applicants are difficult to manage manually to solve this problem recruiter uses machine learning algorithms which makes the screening process a lot easier. Using machine learning candidates' profiles, their experience, personal details, qualifications, and many more. They can match the same qualifications with relevant opportunities. Machine learning does not select the candidate rather than it finds the intangibles and focuses on them this makes the hiring process stronger and good.

2) Compressing the hiring gap – Company's major problem is to fill the vacant positions. Recruiters use machine learning to gather information about the job vacancy. Recruiters have some expectations from the applicants for the job and use the information given by the machine learning to fill the vacancy of the company which is difficult to fill also it is of utmost importance that they deliver the right amount of accuracy for which company will recruit professionals.

3) Job vacancy - If a company wants to hire an engineer, then through machine learning they assess the candidate's performance in the past in the specific role. Using machine learning through specified machine learning algorithms for the unique talent and their qualifications and the unique need of the company. Recruiters also keep in there that the star employee of a company is not compulsory that the applicant is also a star employee of the other company. So, machine learning helps the recruiters to find suitable candidates.

4) Reviewing resumes and social behavior - Sending Resumes is the first part for the applicants in the hiring process and screening is done based on the resumes. Using machine learning information is gathered from resumes their likes, dislikes, and other factors too. This is how machine learning assesses how well the applicants fit the company's culture. Not only their resumes but also their social media and other factors are also assessed for their social behavior using machine learning. Machine learning helps to better understand the hiring patterns and it will help to get identical results to get a unique candidate for the company.

5) Balance's recruiters' risk- Machine learning helps to manage the recruiter's work during the hiring process. This is done through machine learning which takes less amount of time and gives better results, and it gives more time for the company's work.

6) Conducts intuitive search - Another challenge faces, by the recruiters is how to know about the good candidates who are best suited for the job. For which artificial intelligence is used. Machine learning recruitment conducts research based on the data gathered which brings conceptual search, implicit search, and semantic search.

Conceptual search helps to search for a few keywords and implicit search delivery results based on the information which was gathered or assumed by the recruiters. semantic search it is possible to comprehend the searcher's intent.

7) Extensive search process - The main problem with the traditional hiring process is that the recruiters have to search for the candidate which results in missing some of the applicants which are gained by machine learning. Machine learning helps the recruiters so that no candidate is left out during the screening process.

Based on the above points it is very clear that machine learning helps recruiters in recruiting the candidate to a greater extent and it will also help the company in saving resources, money, and time and which gives an incredible positive rate during the recruitment.

9. Regression Analysis

Regression analysis is a statistical method in which we check a relationship between a dependent variable and an independent variable. Dependent variables also predictor variables can be one or more than one. Regression helps in predicting future outcomes such as weather forecasting weather tomorrow's rain will occur or not. This analysis can be done through logistic regression which is a part of regression analysis.

Regression term was named by **FRANCIS GALTON** in the 19th century, but regression analysis use is starting in **1807** by **Legendre** which named as **ordinary least squares** which were one of the methods now to do the regression analysis. IN **1809 Gauss** comes and **Legendre** and Gauss both worked on further study, they found the **Gauss-Markov theorem** which is also a good and popular method nowadays.

Regression analysis is a very popular method and research is going on in this field now. Nowadays regression analysis can be used in many ways such as: -

1) Prediction of weather – very commonly used application I which we predict whether tomorrow is a sunny day or rainy day.

2) Prediction of road accidents due to rash driving – reviewing the past accidents we predict the no of road accidents due to rash driving.

3) Determining the market trends – by viewing and reviewing the past days, we predict how the market is going to work in up comings days or even in a year.

9.1 Why do we use regression analysis?

A continuous variable is always used in regression analysis. Mainly it is used for the prediction. Regression analysis is greatly applied in the application of machine learning and data science. Some real-world examples are weather forecasting, market trends, risk assessment, score prediction, etc. therefore regression analysis is such an important tool that we can apply in machine learning and data science. There are some reasons why we use so much regression analysis is given below:

1) To analyze a wide variety of relationships-

Variables can be of different types such as categorical variables, numeric variables, polynomial, binary variables, etc. so finding the relationship between them is important. Regression analysis helps in finding the relationship between the different types of datasets.

2) To control independent variables –

When a relationship is created between target and predictor variables then changes occur in the variables which were controlled by regression analysis.

When the regression analysis is done, we freeze the role of predictor variables. For this, we need to minimize the confounding variables which means that to minimize the biases in the model then we get to know about every variable separately and it does not affect by any other variables.

During regression analysis, we control the other independent variables by putting them in the model. During that time, some relevant variables can be misleading to the results. Omitting variables causes the model to be uncontrolled and always be biased toward the variable which is not present in the model.

3) To interpret regression output –

In regression analysis, first, we fit the model and then we verify the model that the model we created is a good model or not. for this, we look at the regression coefficients and p-values. If every variable has less than a 0.05 p-value, then the value is significant.

p-value helps you to determine whether the relationship which is works with the sample also works with the population too.

4) To obtain trustworthy regression results.

If we want a trustworthy model, then we must check for the following outcomes which are:

a) If we are not including all the variables in the model, then the model is biased.

b) The correlation between two independent variables is called multicollinearity but it can be in the right order. If multicollinearity exceeds then it is a problem and creates wrong results.

c) We must check the residual plots. Residual plots are graph that shows the residual value and the independent variables. Residual value checks how much a regression line misses a data point. We must be sure that the model does not leave any data point otherwise the model is not fit adequately.

9.2 Terminologies related to Regression Analysis

Dependent variable – this is one of the main factors in the regression analysis. This is a factor that we want to predict or understand is called the dependent variable.

Independent variable – the factors which are used to predict the values of the dependent variable are known as independent variables. They are also called predictors.

Outliers – it is an observation that is very high value or very low value from other observation is called outliers. Outliers can make our results weak. We should not include outliers in our results.

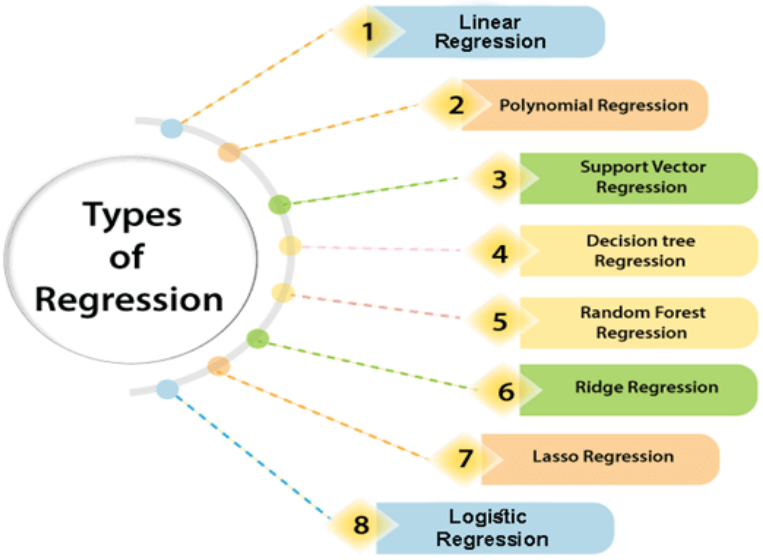
Multicollinearity – when independent variables are highly correlated variables with each other then it is called multicollinearity. It should not include in the dataset otherwise it will affect the whole dataset while ranking the most affecting variables.

Underfitting and Overfitting - when the algorithm is well suited for training data but not for testing data then it is known as overfitting.

But when the working algorithm is well suited for testing data but not suited for training data then it is known as underfitting.

9.3 Types of regression analysis

There are mainly seven kinds of regression analysis which are: -

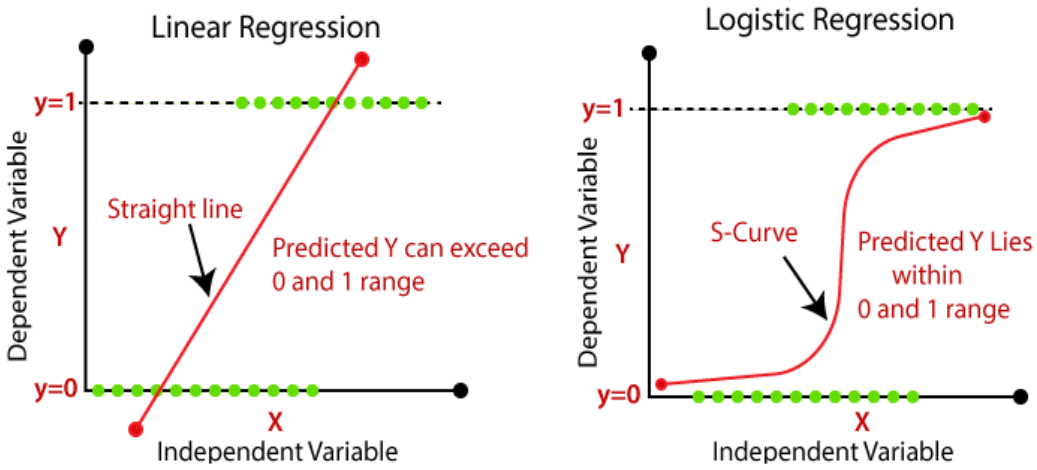


1) Linear Regression
2) Logistic Regression
3) Polynomial Regression
4) Support Vector Regression
5) Ridge Regression
6) Lasso Regression
7) Decision Tree Regression
8) Random Forest Regression

10. Difference between linear regression and logistic regression

Linear regression	**Logistic regression**
In linear regression, we predict the value of continuous variables.	In logistic regression, we predict the value of categorical variables.
In linear regression, it is used to solve regression problems.	In logistic regression, it is used to solve classification problems.
In linear regression, we will fund the best fit line through which we can predict the output.	In logistic regression, we will form the S- curve through which we can classify the samples.

In linear regression, the least-square estimation method is used to show the accuracy of the model.	In logistic regression, the maximum likelihood estimation method is used to show the accuracy of the model.
In linear regression, the output of linear regression will be a continuous variable such as age, price, etc.	In logistic regression, the output of logistic regression will be a categorical variable such as 0 or 1, yes or no, etc.
In linear regression, there may be collinearity between the independent variables.	In logistic regression, there should not be collinearity between the independent variables.
In linear regression, the relationship between the dependent variables and the independent variables should be linear.	In logistic regression, the relationship between the dependent variables and the independent variables doesn't have to be linear.
Linear regression is used to predict continuous dependent variables using a given set of independent variables.	Logistic regression is used to predict the categorical dependent variables using a given set of independent variables.
Examples are score prediction which is a very common example.	Examples are spam detection which is shown on every mobile phone in every mail have.



This is the graph showing the linear regression and logistic regression in which

The linear regression has a straight line in which the value of y can go beyond the 0 and 1 range. It also has outliers. It cannot give probabilistic values.

Whereas in logistic regression the curve will be in S-shaped form and the value cannot go beyond the 0 and 1 range. It gives probabilistic values.

11. Classification algorithm

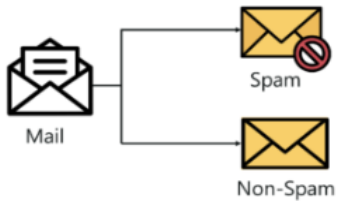
Classification in machine learning is a supervised machine learning algorithm in which the machine learns from the data given to it and makes the classification. It is a process through which categorizes a given set of data into classes. It can perform in structured and unstructured data. The procedure starts with predicting the points of class in the given data. Classes are also called target, label, and categories.

In the classification algorithm, predictive modeling approximates the mapping function from input variables to discrete output variables. The main objective of the classification algorithm is to find in which class the new data fall.

For example

Spamming the mail is very common in the mail. Spam detection in the mail is an example of a classification algorithm because whether the mail is coming to the person is spam or not. This is the binary classification which has only two classes. In this case, first, the training data needs to understand how the training data is related to the class. After this, it is predicted on the testing data whether the mail is spam or not.

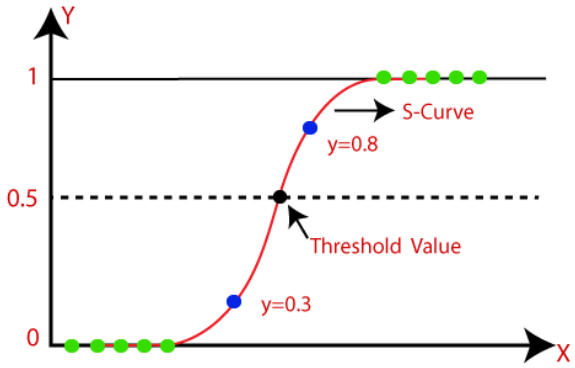
Since classification is a type of supervised machine learning classes are provided with input data and they also have the desired output.



12. Logistic Regression

Logistic regression is one of the most popular machine learning algorithms under the supervised machine learning algorithm. it is used to solve classification problems. Logistic regression is used for predicting the dependent variable using a given set of independent variables. It is a classification problem where the target variable is categorical. It can be either yes or no, true or false, type A type B type C, etc. Logistic regression gives the probabilistic values which lie between 0 and 1. Since the value lies between 0 to 1 so line which was formed on the graph is in ‘S’ form. The curve tells us whether the person is suffering from obesity or not, whether an email is a spam or not, whether cells are cancerous or not, etc. there are many examples, and the medical department uses logistic regression to a greater extent. The curve indicates the likelihood function. Logistic regression is a significant machine learning algorithm as it gives probabilities and classifies new data using discrete and continuous datasets. Logistic regression is used to classify the observations using different types of data and then it determines the most effective ways used for classification. The below image is shown for the logistic function. Logistic regression uses predictive modeling as regression therefore, it is called logistic regression but it is used to classify samples, therefore, it comes under the classification algorithms.

12.1 Sigmoid Function



1) The sigmoid function is also known as the logistic function.

2) The sigmoid function is a function used to map from predictive values to probabilities.

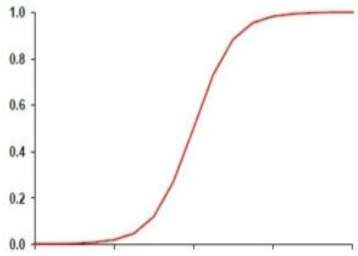
3) It can do a map for any value between the values 0 and 1.

4) The value of the sigmoid function should lie between 0 and 1 which cannot go beyond this value therefore it forms the S form. The s-form is called a sigmoid function or the logistic function.

5) In logistic regression, we use the threshold value which also lies between 0 to 1. If the p-value is less than the threshold value, then it tends to 0 or if the p-value is greater than the threshold then it tends to 1. Mainly we take the threshold value to be 0.5.

12.2 Types of logistic regression

There are three types of logistic regression which are as follows



1) Binomial – in binomial logistic regression there can be only two possible dependent variables such as 0 or 1, true or false, yes or no, etc.

2) Multinomial – in multinomial logistic regression there can be three or three unordered types of dependent variables such as “cats”,” dogs”, and” mice”.

3) Ordinal – in ordinal logistic regression there can be three or more three ordered types of dependent variables such as “low”,” medium”, and” high”.

12.3 Assumptions of Logistic Regression

There are two assumptions of logistic regression which are as follows

1) The dependent variable is categorical.

2) The independent variable should not have multi-collinearity which means that the independent variables must be independent of each other.

3) All the variables in the model must be meaningful that is unnecessary variables are not included in the model.

4) The sample size should be large.

12.4 Logistic regression Equation

The logistic regression equation is the different linear regression equation. the mathematical equation is as follows

1) The equation of the straight line is as follows

$$y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

2) In logistic regression, y can be between 0 and 1 only. For this, we have to divide the equation by (1-y)

$$\frac{y}{1-y}$$; 0 for y= 0, and infinity for y=1

3) But we need a range between –[infinity] to +[infinity], then the log to both sides. It will become

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

This is the final equation of the logistic regression.

12.5 Application of Logistic Regression

Logistic regression is a very commonly used method in machine learning. Medical areas use logistic regression very much. Some of the main applications are as follows

1) Determining a heart attack
   Medical researchers use medical data to understand the relationship between the predictor variables that the person will have a heart attack or not. Based on age, sex, diet, fitness any more factors researchers analyze the changes and how it affects the person for having a heart attack.

2) Understanding the possibility of getting admission into a university
   For college admission person's marks such as CGPA, GMAT, etc. are very important. Scores will determine whether the person is getting into university or not. Therefore, logistic regression is used for determining whether a person is getting selected or not by showing the relationship between the predictor variables and the probability of getting selected.

3) Gmail and other inboxes find spam Emails
   One of the most common examples which person has seen every day. Spam mail in which all the unnecessary mail come which we do not want. Filtering in spamming is done through logistic regression by checking whether the mail is promotional or not. if the mail is promotional then the mail sends to spam otherwise not.

13. Confusion Matrix

The confusion matrix is a matrix used to decide the performance of the classification models from the testing data. It can only be decided if the true value values are known in the testing data. The terminologies are confusing, but the matrix is easy to understand. Therefore, it is called confusing, and since it is the matrix, so it is called confusion matrix. Since it shows errors in the model which was created through training data in the form of the matrix, therefore it is also called an error matrix.

Some features of the confusion matrix are following as

1) For the prediction of classes of classifiers, the matrix should be of size 2x2, 3x3, 4x4, etc.

2) The table is divided into two parts which are the predicted values and the actual values along with the number of predictions.

3) The model predicts the predicted values.

4) Actual values are true values that were given for the observations.

It is shown in the given table

Total predictions	Actual: No	Actual: Yes
Predicted: No	True Negative	False Positive
Predicted: Yes	False Negative	True Positive

In the above table there are 4 cases which are as follows:

1) True positive –
   The predicted value of a model, was Yes, and the actual value or true value was also No.

2) True negative –
   The predicted value of the model is No, and the actual value is also No.

3) False Negative –
   The predicted value of the model is No, but the actual value is Yes, therefore it is called a TYPE-II error.

4) False-positive –
   The predicted value of the model is Yes, but the actual value is No, therefore it is called a TYPE-I error.

13.1 Need for the Confusion Matrix

Without there being any need, we are not going to use a particular thing. There are some ways through which we can tell about the need for a confusion matrix which are as follows:

1) it calculates the performance of classification models when the predictions are made on the testing data and tells us how good a model is.
2) It not only talks about the errors but also talks about the type of error whether it is a type-I error or type-II error.
3) From the confusion matrix, we can calculate different parameters of the model, such as accuracy, precision, etc.

13.2 Calculations using a Confusion Matrix

We can calculate various calculations using a confusion matrix.

These calculations are given below

1) **Classification accuracy** – it is one of the most important parameters to decide the accuracy of the classification problems. It tells us how frequently the model predicts the correct output. It is calculated as the ratio of the number of correct predictions made by the classifier to all the predictions made by the classifier.

   $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP}$$

2) **Misclassification rate** – it is known as the error rate; it tells us how frequently our model gives the wrong predictions. The value of the error rate can be calculated as the ratio of the number of incorrect predictions to all number of predictions made by classifiers. It can be calculated as follows:

   $$\text{Error rate} = \frac{FP+FN}{TP+FP+FN+TN}$$

3) **Precision** – it is the number of correct outputs given by the model, or all positive values correctly predicted by the model, or how many of them are true. It can be calculated by the formula which is as follows

   $$\text{Precision} = \frac{TP}{TP+FP}$$

4) **Recall** – it tells us that out of total positive classes, how our model correctly. The recall must be as high as possible.

   $$\text{Recall} = \frac{TP}{TP+FN}$$

5) **F-measure** - if two models have low precision and high recall or vice-versa, therefore it is difficult to calculate. So, we calculate the f-score. F-score helps us to calculate the recall and precision at the same time. It can be maximum if a recall is equal to the precision. It can be calculated as follows

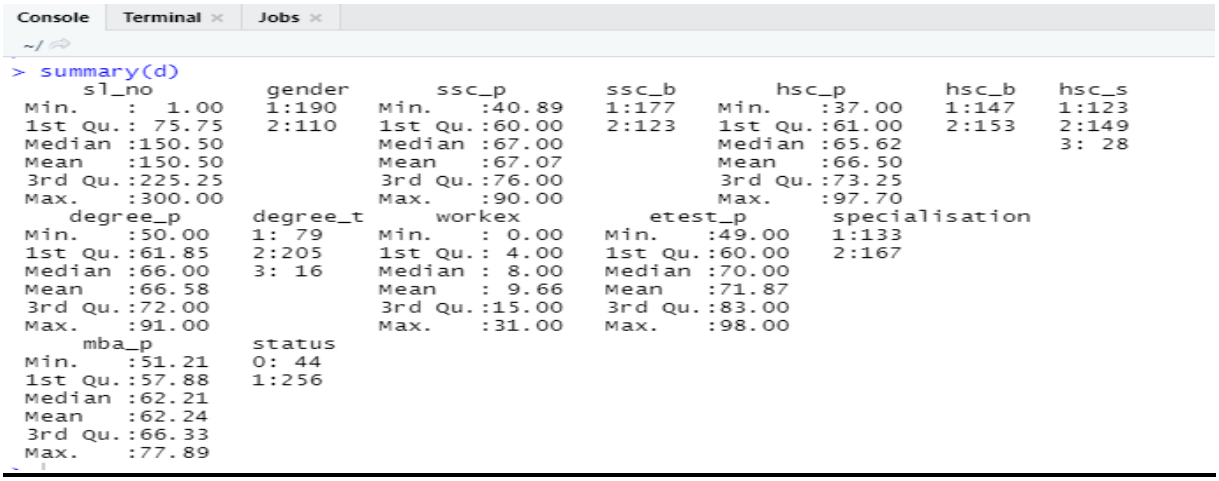
   $$\text{F-measure} = \frac{2*recall*precision}{Recall + precision}$$

6) **Null error rate** – it tells us how frequently our model will be incorrect if it always predicted the majority class. As per the accuracy paradox, it is said that "the best classifier has a higher error rate than null error rate".

7) **ROC curve** – it is the graph showing all possible thresholds for displaying classifiers' performance. It is plotted between the true positive rate (on the y-axis) and the false positive rate (on the x-axis).

14. ANALYSIS

1) Table 1



Console | Terminal | Jobs

~/

```
> summary(d)
      sl_no          gender      ssc_p           ssc_b       hsc_p           hsc_b     hsc_s
 Min.   :  1.00   1:190   Min.   :40.89   1:177   Min.   :37.00   1:147   1:123
 1st Qu.: 75.75   2:110   1st Qu.:60.00   2:123   1st Qu.:61.00   2:153   2:149
 Median :150.50           Median :67.00           Median :65.62           3: 28
 Mean   :150.50           Mean   :67.07           Mean   :66.50
 3rd Qu.:225.25           3rd Qu.:76.00           3rd Qu.:73.25
 Max.   :300.00           Max.   :90.00           Max.   :97.70
    degree_p     degree_t      workex         etest_p       specialisation
 Min.   :50.00   1: 79   Min.   : 0.00   Min.   :49.00   1:133
 1st Qu.:61.85   2:205   1st Qu.: 4.00   1st Qu.:60.00   2:167
 Median :66.00   3: 16   Median : 8.00   Median :70.00
 Mean   :66.58           Mean   : 9.66   Mean   :71.87
 3rd Qu.:72.00           3rd Qu.:15.00   3rd Qu.:83.00
 Max.   :91.00           Max.   :31.00   Max.   :98.00
     mba_p       status
 Min.   :51.21   0: 44
 1st Qu.:57.88   1:256
 Median :62.21
 Mean   :62.24
 3rd Qu.:66.33
 Max.   :77.89
```

This table shows the summary of every variable. The numerical variable has its minimum, maximum, median, 1st quartile and 3rd quartile whereas the categorical variable is showing classes of a particular set.

2) Table 2

```
Console   Terminal   Jobs
~/

Call:
glm(formula = status ~ ssc_b + ssc_p + hsc_b + hsc_p + hsc_s +
    degree_t + degree_p + mba_p + specialisation + etest_p +
    workex, family = binomial, data = training_set)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6124   0.2133   0.3661   0.5306   1.8734

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    8.8473399  2.9246854   3.025  0.00249 **
ssc_b2         1.0746729  0.6013429   1.787  0.07392 .
ssc_p          0.0033517  0.0213382   0.157  0.87519
hsc_b2        -0.5093343  0.5450385  -0.934  0.35005
hsc_p          0.0005541  0.0251616   0.022  0.98243
hsc_s2        -0.2795573  0.5324357  -0.525  0.59955
hsc_s3        -0.6139939  0.7281777  -0.843  0.39912
degree_t2      0.8741330  0.5194364   1.683  0.09240 .
degree_t3      0.0556291  0.9088404   0.061  0.95119
degree_p      -0.0097398  0.0374730  -0.260  0.79493
mba_p         -0.0840924  0.0425790  -1.975  0.04827 *
specialisation2 -0.0112439  0.4747857  -0.024  0.98111
etest_p       -0.0053105  0.0166132  -0.320  0.74923
workex        -0.1269230  0.0285739  -4.442 8.92e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 187.60  on 224  degrees of freedom
Residual deviance: 149.66  on 211  degrees of freedom
AIC: 177.66

Number of Fisher Scoring iterations: 5
```

This table is showing the logistic model of the recruitment data. We have seen that this model has the lowest AIC value which means that this is the best model. It shows that work (work experience), mba_p(MBA marks), and degree_p (degree marks) has the most significant variable which means that they will highly affect the recruitment process. The summary of the model is shown above. This model is based on the training set which is created for building a model.

3) Table 3

```
Console   Terminal   Jobs
~/
> prob_pred
        3         5         6        10        13        16        20        22
0.9445368 0.7200435 0.9689417 0.9843058 0.8757782 0.6592163 0.6791735 0.9261143
       25        26        29        31        34        40        42        44
0.6348227 0.7823439 0.9615638 0.9723794 0.7896528 0.8257246 0.9554233 0.9398300
       47        64        67        69        75        76        83        85
0.9481467 0.9860726 0.7054606 0.9380911 0.9465469 0.8349518 0.9357553 0.7079839
       86        87        88        90       103       107       108       109
0.6725613 0.9727564 0.9426411 0.9127881 0.9677975 0.9440679 0.8836825 0.9748885
      110       118       127       129       130       134       137       142
0.3637008 0.8711806 0.8905411 0.7405269 0.3597545 0.7228893 0.4743718 0.9386297
      152       158       160       163       166       167       174       180
0.9427420 0.9442290 0.9586992 0.5875820 0.1767360 0.9803592 0.9767044 0.4733255
      181       200       206       209       214       216       224       226
0.8136725 0.9860547 0.8893034 0.7444831 0.9588301 0.9583451 0.9661685 0.9119330
      229       231       239       242       245       256       259       260
0.5754752 0.9663145 0.8826683 0.8145208 0.9016260 0.1969036 0.7921372 0.7756601
      261       263       264       272       281       283       291       292
0.9380969 0.9636071 0.9816006 0.2413427 0.9364691 0.9906084 0.9097076 0.8392920
      293       294       300
0.9658505 0.7387527 0.9052846
```

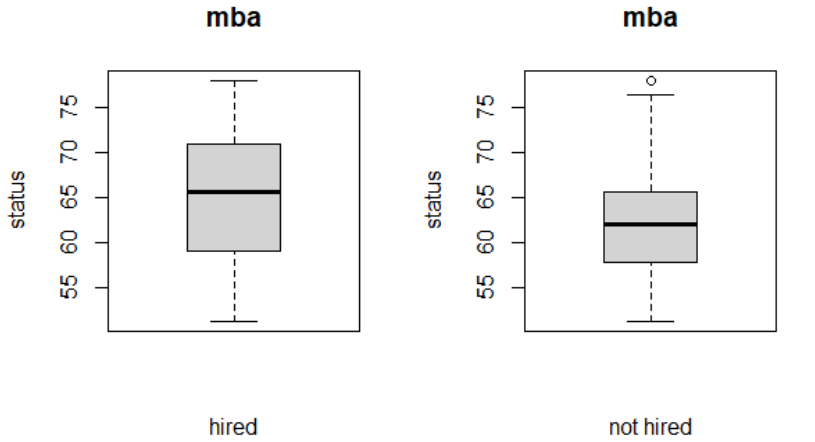
This table is showing the probabilistic value prediction which is done on the testing data. It is showing the value between 0 to 1.

4) Table 4

```
Console  Terminal  Jobs
~/
> y_pred
  3   5   6  10  13  16  20  22  25  26  29  31  34  40  42  44  47  64  67  69  75
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
 76  83  85  86  87  88  90 103 107 108 109 110 118 127 129 130 134 137 142 152 158
  1   1   1   1   1   1   1   1   1   1   1   0   1   1   1   0   1   0   1   1   1
160 163 166 167 174 180 181 200 206 209 214 216 224 226 229 231 239 242 245 256 259
  1   1   0   1   1   0   1   1   1   1   1   1   1   1   1   1   1   1   1   0   1
260 261 263 264 272 281 283 291 292 293 294 300
  1   1   1   1   0   1   1   1   1   1   1   1
> summary(y_pred)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  1.0000  1.0000  0.9067  1.0000  1.0000
```

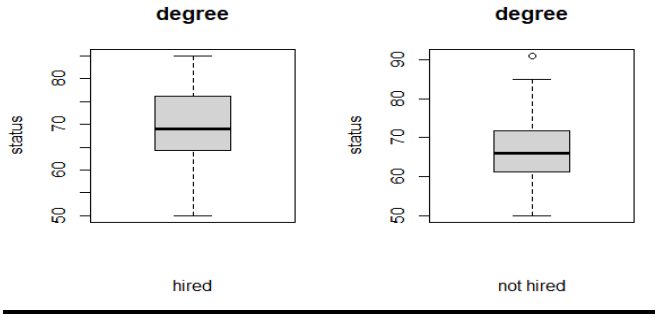
This table shows the prediction of the given data which is firstly showing the probabilistic value. By using the threshold value which is taken as 0.5. it will answer 0 or 1. If the p-value is less than 0.5 then it will give 0 otherwise it will give a value of 1.

1) Figure 1



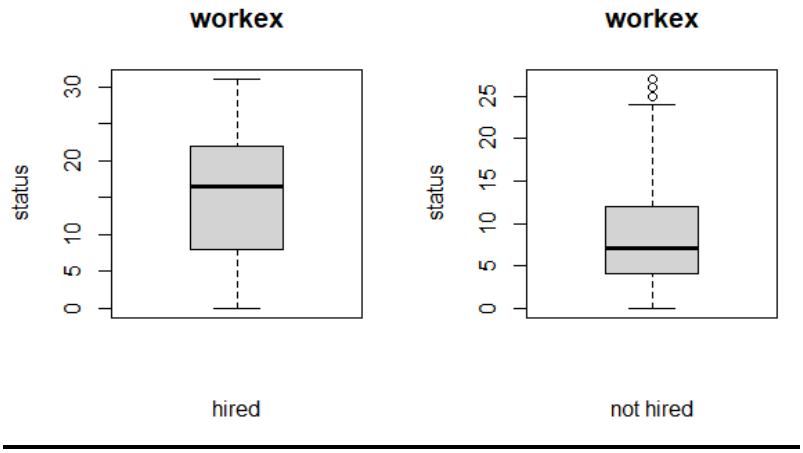
MBA marks also affect the recruitment process. This figure is showing that a hired candidate has an average no of marks and not hired candidates have mainly less than average marks

2) Figure 2



It is showing the marks on graduation. This figure is showing that a hired candidate has an average number of marks in graduation whereas a not hired candidate has lower than average marks in graduation.

3) Figure 3



The figure is showing the number of months experienced by candidates before joining the post-graduation course. From the model, we can see that it has the highest influence on the model that if the person has work experience, then the candidate has an advantage, and it gives weightage in the hiring process.

The figure is showing that hired candidates have average work experience and not hired candidates have extraordinarily little experience in work.

15. Results

In this project, online recruitment system has revolutionized the offline recruitment system. Slowly by slowly in some ways there is a change in the recruitment process which highly affects the traditional method. From the data we have collected we have clearly seen that a company is hiring a fresher by its qualifications specially by their work experience and their MBA marks for their assistant manager post. As we have seen that for acquiring any job, we have to make our educational qualifications looks good. A student is getting hired or not is mostly depends on the academics for the first time. But apart from these qualifications students who is getting graduation marks was also important for some candidates from the process of recruitment. Therefore, we can see that recruitment system is somehow changing and using machine learning makes easy way for the recruiters and the candidates so that it will makes us life easier.

16. Conclusion

We conclude that machine learning helps the recruitment process in many ways to ease the work of a recruiter or the candidate who is filling the post in a company or finding a job in the company. As it will also tell us work experience is most important for the post of assistant manager. Experience and marks in graduation and MBA play a crucial part in getting posted by any company. Therefore, a good work experience will get too easy for a candidate to get a job in a reputed company. We also see that hired candidates have an average in marks or inexperience. They neither are too good nor too bad. Hiring can be difficult to make our process fast we have to use machine learning. Our traditional system of hiring candidates is outdated and needs to be replaced as it will waste a huge amount of time, money, and mainly services that will use somewhere where it is also needed.

17. References

https://en.wikipedia.org/

https://machinelearningmastery.com/

https://www.tutorialspoint.com/

https://www.careeraddict.com/

https://recruiteze.com/

https://www.candidatemanager.net/

https://get.fountain.com/

https://www.topechelon.com

https://www.geeksforgeeks.org/

https://books.google.co.in/

https://science.sciencemag.org/

https://www.adacovid19.org/

https://www.sciencedirect.com/

https://www.analyticsvidhya.com/

https://www.edureka.co/

https://www.digitalocean.com/

Introduction to machine learning by Nils J Nilsson

https://www.udemy.com

https://www.investopedia.com/

https://www.techfunnel.com/

https://www.sciencedirect.com/

https://www.digitalvidya.com/

https://marketbusinessnews.com/

https://www.tutorialspoint.com/

https://in.springboard.com/

https://www.javatpoint.com/

https://towardsdatascience.com/

https://in.springboard.com/

https://www.besanttechnologies.com/

https://www.stat.cmu.edu/

https://stackoverflow.com

https://www.datamentor.io/

https://www.forbes.com/