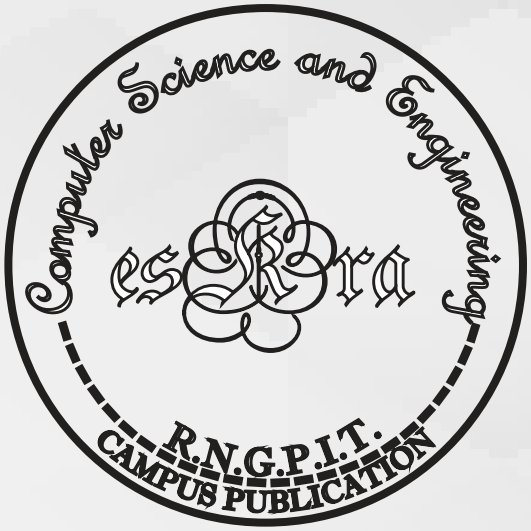
ESKRA

THE FUTURE IS HERE



Circulation:- ESKRA is published frequently by the students of Computer Science and Engineering department of R. N. G. Patel Institute of Technology, Bardoli-Navsari road, At:Isroli, Taluka:Bardoli-394620, District:Surat.

Phone: +91-9427527398, +91-9033707334, +91-9510427118 e-mail: eskra.cse@rngpit.ac.in

Editorial:- Unless otherwise stated articles,as well as images and description, reflect author's or firms opinion. Inclusion in ESKRA does not necessarily constitute endorsement by ESKRA or students of Computer Science and Engineering department of R.N.G.P.I.T. All submissions are subject to editing for style, clarity,and space.

Reserve rights and reprint permission :-

Message from Editor-in-Chief

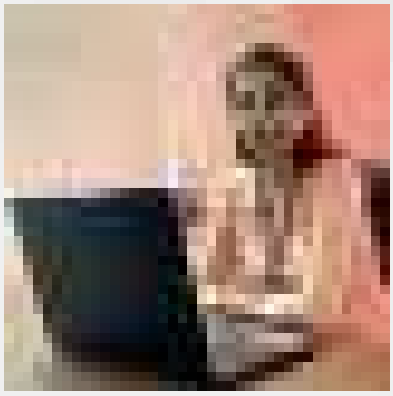
Dr. Madhavi B. Desai
(HoD & Associate Prof., CSE, RNGPIT)
(Editor in Chief)

Being a part of computer department, we are always eager to know things that are related to it. We always enthralled by technologies. Hence, the mere fact of knowing about Machine Learning and Artificial Intelligence makes us feel aroused. We have published peer-reviewed articles in this magazine that give us an introduction to AI, ML and Data Science, some basics of using different applications, and an attempt to justify a few uncertain questions whose answers vary depending on people's perspective.

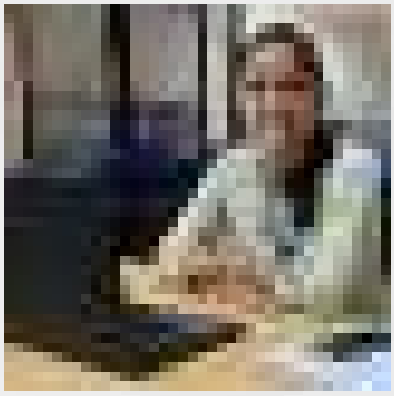
Our goal is to be at the forefront of disseminating scientific knowledge and impactful discoveries to academics, and the general public around the world. "FUTURE IS HERE" take us to a journey of things which are related to AI, ML and Data Science and put an eagerness into reader's mind to delve deeper. Each article in this magazine is consigned with the intention of bestowing a trove of knowledge on each reader.

It's very common these days to come across these terms - data science, artificial intelligence, machine learning and much more. But what exactly do these soundbites mean? And why should you care about either one? To the best of our ability, we have attempted to answer these questions in this magazine.

Editorial



Ms. Bhavini R. Bhatt
(Asst. Professor CSE, RNGPIT)
(Faculty Advisor)



Ms. Dharmishtha R. Chaudhari
(Asst. Professor CSE, RNGPIT)
(Faculty Advisor)

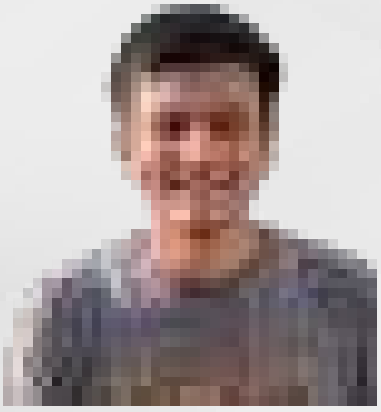
Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Machine learning is the study of computer algorithms that improve themselves automatically over time. It is considered a subset of artificial learning. Machine learning is the process by which computers learn how to perform tasks without being explicitly programmed to do so. It's a vast area with unbridled prospects and much more to discover. Data science refers to the process of extraction of useful insights from data.

For a simpler view on the relation between these technologies, artificial intelligence is applied based on machine learning. And machine learning is a part of data science that draws features from algorithms and statistics to work on the data extracted from and produced by multiple resources. Thus, you can say data science merges together a bunch of algorithms obtained from machine learning to develop a solution, and during the process, lots of ideas from traditional domain expertise, statistics and mathematics are borrowed.

We hope that this magazine will assist you in augmenting your thinking and raising your level of imagination and acceptance to new heights.

About Authors

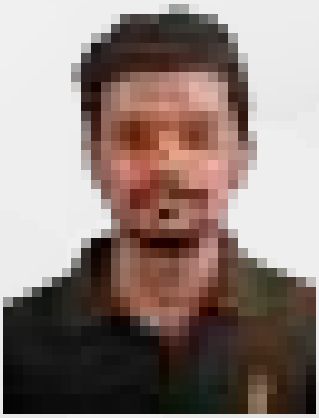
Suraj Parmar



Suraj is a ALUMNI student of Computer Science and Engineering department of R.N.G.P.I.T. He is a passionate technological writer from last few years and wrote some articles on certain topics during his engineering. He is highly skilled and educated. He provided his knowledge about gradient descent, autoencoders and machine learning through his writings to outshine in this magazine.

Articles:- Gradient Descent, Machine Learning in the Browser using TensorFlow.js, AutoEncoders-Denoising Understanding, Regression using Sklearn.

Visarg Desai



Visarg wrote about the significance of p value in machine learning. He is a ALUMNI student of Computer Science and Engineering department of R.N.G.P.I.T . He put down this article to share his understanding on a particular topic in machine learning with his creative and easy to remember writing skills.

Articles:- Data Science

Nisarg Mahyavanshi

Nisarg is a FOURTH year student of Computer Science and Engineering department of R.N.G.P.I.T. Using his unique, hyperbole and user friendly technique of writing, he has shared his experience and knowledge of Artificial intelligence so that readers can gain knowledge regarding the same.

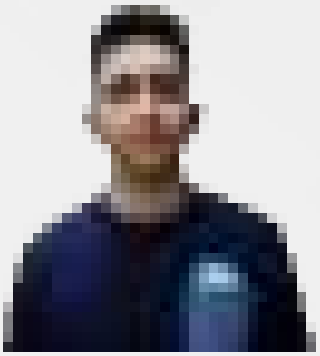
About Authors

Shreya Lad

Shreya is a FOURTH year student of Computer Science and Engineering department of R.N.G.P.I.T. Through her article she not only wants to show her knowledge but also wants to lend help to those students who have interest in the field of machine learning. She wrote an article on beginners in data analytics in machine learning in this magazine.

Articles:- Beginnners in Data Analytics using Machine Learning..?

Arjav Desai



Arjav is a FOURTH year student of R.N.G.P.I.T. He wants to present something which is from the core of his abilities, so that the reader will enjoy his writings and knowledge regarding Data Science.

Articles:- AI to replace Doctors: A trivial untruth

Contents

1. Symbolic Artificial Intelligence

- Nisarg Mahyavanshi
(FOURTH year, CSE, R.N.G.P.I.T)

You :- So what is the difference between AI and ML
Internet :- ML is a subfield of AI
You :- okay then what comes in AI but not in ML
Internet :- Vision, NLP, Robotics ….etc
You :- But doesn't that require ML algorithms??
Internet :- Yes
You :- *Visible Confusion*…….

This is generally the scenario when you come across an article/blog/video on the Internet which not only confuses you more but makes you question the information you already possess about AI. It is hard to find such articles that fully answer the question and make your life less miserable. So here we'll try to discuss it way simple and see it bit by bit but you have to bear some history (*don't worry we'll keep it short*). So what's all this fuss about, well you see in the present time Machine Learning has advanced so much that when someone new enters in the field of AI all they hear is ML, ML, ML… so much that difference between them gets fuzzy, and while it is true that ML is awesome, there are other mind-blowing fields too. Have you ever heard of "SYMBOLIC AI" or "GOFAI" or "GOFR" (Let me tell you this technology is powerful enough to create bare bones of "CHITI THE ROBOT" or "JARVIS" which ever you prefer). While ML is the tech that solves subproblems like NLP, vision, robotics, autonomous vehicles…and much more.
Let start with the basics (here comes the history) -

What is Artificial Intellgence?

It's a term coined in an article of the 1950s when a brilliant (& lazy I guess) mind thought of a technology that can make decisions on its own (just like a human)with the data provided to it and the very first program based on AI was written in 1951. (Machine Learning wasn't even born yet) This technology progressed a lot after 1980 but soon this giant fell. This period wasn't good for AI and many even called it as "AI winter"

What is Machine Learning?

It's a technology that is used to search for a pattern in the given set of data and then make predictions for new but relevant data from the pattern it has found. (Yes, you might be confused but bare with me it'll make sense after some time). It was first coined in 1959 (nearly a decade after AI) and gradually started its dominance in the field of data analytics. Also its most popular approach to AI.

So, What is the Difference?

The Main Goal of an AI model is to mimic human intelligence as closely as possible or beyond it while on the other hand The main Goal of an ML model is more towards recognizing a pattern in the data set. Yes an ML model that can not make decisions is still an ML model if it can recognize the pattern or structure of data. But what's the use if it can not make decisions right??

Let's take an example -

Suppose you are a foodie and today you want to enjoy your dinner at a nice restaurant which you haven't visited yet so you ask your google assistant (or some similar bot) to show you top restaurant in the neighbourhood. It will show you a list of restaurants which are famous around you. Did it decided for you??, No cause final decision still depends on you. So can we say it is an ML model, Yes because it still made the list based on various parameters like how many people liked it, does it maintain cleanliness, is service and food good there, is it well rated, etc. Is it an AI, well answer it yourself does it mimic a human brain somehow no right. But now you may argue that a google assistant is an AI well yes because it does much more than selecting a restaurant for you which differentiates it from a normal ML model. (Note we aren't referring google assistant whole as an ML model we are referring to that particular feature which shows the list of restaurants. Google assistant is, in fact, an AI which comes under the category of "Narrow Artificial Intelligence"). So yes an ML model can exist without becoming an AI. But is it's reverse possible, Can an AI exists without an ML model. Before continuing try to think about an example of it. Well, I hope you made that effort and might have found it but still I'm gonna write about it anyway so hang in there. For those who weren't able to figure it out the answer is*drum rolls*. Affirmative (*Of course right bummer*). Well, yes, it is possible to build an AI without ML but how right, The whole decision making process needs to find a pattern in the data on which decisions can be made which is what ML does so how can we build an AI that can decide without it. Well answer to that question is "Symbolic Artificial Intelligence" (check here for more information) Hero Of this story, whose legends are spreading in the valley till this day (*Okay enough*).

Symbolic Artificial Intelligence

So, what is it and how it is different than ML, well whole concept of Symbolic Artificial Intelligence (also known as “Good Old Fashion Artificial Intelligence” or “GOFAI” *believe I’m no making stuff up this is what it’s called*) revolves around the Symbols and their relation. Well, what are those?? A Symbol can be referred to as a real-world entity(constants) and a Relation(predicates) defines how these Symbols relate to each other.

Let’s clear it more with an example, suppose you defined a relation like between A, B, C entities:-
Father(A) →B //Father of A is B
Father(B) →C //Father of B is C
Grandfather(A) →C //Grandfather of A is C
Now you introduce a new entity D like Sister(A) →D //Sister of A is D
Than by logic we can say that Father (D) →B and Grandfather(D) →C
That’s it.

Figure-1.1 Symbolic Artificial Intelligence [1]

It’s that simple (well until more complicated relations are made and it’s very helpful if you are familiar with terms predicates and constants if not than focus on the above analogy). This is how the first AI model was invented. If you are interested watch a GOFAI playing GridWorld. The advantages are below:-
-Unlike ML, here the whole prediction process is Interpretable.
-Decides with the concept level through reasoning.
-Well established algorithms developed back then.

If it's this advantageous than why does this has to pass through such a rough phase like AI winter. It's actually due to its one disadvantage i.e. All knowledge and Base relations need to be handcrafted by the experts.

Its handcrafted with more data it gets hard to construct such models. This is the reason why ML took over this position since once ML models are constructed they can train automatically without any human dependencies and it is advantageous in modern days since there are lots and lots of data. But still, after this rough phase, it's not yet forgotten it is still used by some companies and also used in the research field of "General Artificial Intelligence" which aims to create a human-like Artificial Intelligence like Terminator (*just kidding… or am I*).

Figure-1.2 Eliza [2]

Say Hi to Eliza

Summary :- Saw some history, gave general definition to AI and ML, saw how they can exist independently and an alternative way to approach AI.

Conclusion :- Artificial Intelligence is a huge field and frankly we have just peeped through a hole. Machine learning is Cool (it is awesome actually) but it’s not the only field so don’t narrow down AI only to ML.

Image Source:-

[1] https://miro.medium.com/proxy/1*huTsMf3UW0CBDB-fT55I_g.jpeg
[2] https://miro.medium.com/max/776/1*8o08WKTA-DMSS1HVQqawTA.jpeg

2.AI to replace Doctors: A trivial untruth

- Arjav Desai
(FOURTH year, CSE, R.N.G.P.I.T)

Figure-2.1 AI as Doctors [1]

Artificial Intelligence (AI) has been on the rise for the last decade. According to Wikipedia the definition of AI is “a system’s ability to correctly interpret external data, to learn from such data, and to use those learning to achieve specific goals and tasks through flexible adaptation”. The primary aim of health related AI applications is to analyse relationships between prevention or treatment techniques and patient outcomes. The potential of AI to support various aspects of health sector represents a major leap forward that can result in prevention strategies and treatment approaches that are much more efficient and effective, including the reduction of medical care costs. Nonetheless, some doctors believe that, although AI can perform a variety of tasks, it cannot replace human interaction.

So, where is AI being used in the healthcare sector?

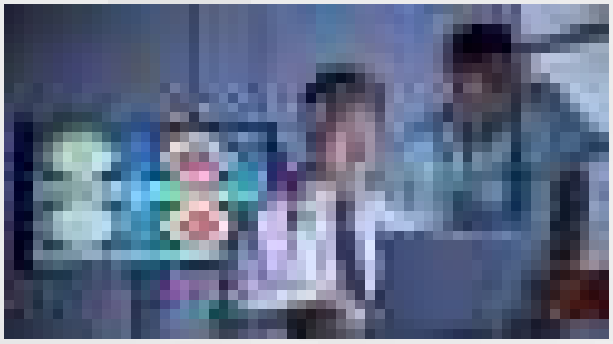


Figure-2.2 Remote Surgery [2]

1)Remote Surgery

Remote surgery is a method, which provides the doctor to perform surgery remotely through a robotic surgical system from anywhere in the world. It promises to allow the expertise of specialized surgeons to be available to patients worldwide without the need for patients to travel beyond their local hospital. But there is a caveat the cost of such systems is approximately near $1 million and the operational costs ranging from $150k to $200k. Before its acceptance on a broader scale, many issues will need to be resolved. For example, established clinical protocols, training, and global compatibility of equipment must be developed. Another technological limitation is the risk of interference with communications (hacking). Before its acceptance on a broader scale, many issues will need to be resolved. For example, established clinical protocols, training, and global compatibility of equipment must be developed. Another technological limitation is the risk of interference with communications (hacking).

2)Virtual Nurse

Virtual nurses are virtual assistants designed to perform the duties of an orthodox nurse. They carry out routine tasks such as monitoring the patients' stats and are programmed to provide a personalised experience to the patients. They cannot, however, provide empathy in the same way that a regular nurse can.

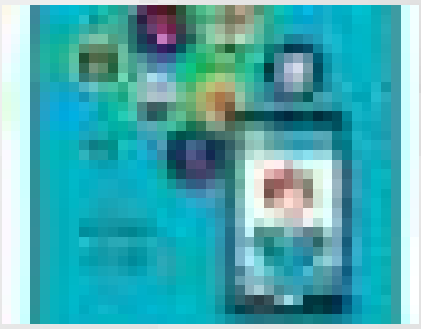


Figure-2.3 Virtual Nurse [3]

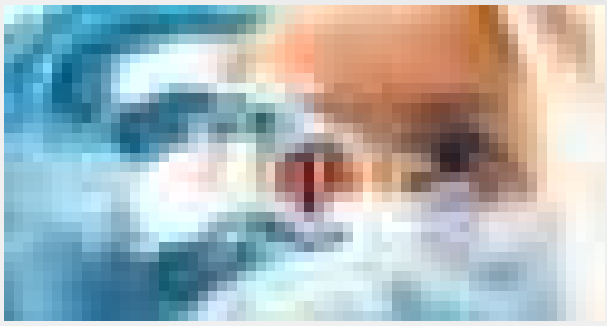


Figure-2.4 Discovering new drugs [4]

3)Discovering new drugs

AI is now being used by big pharma companies to discover new drugs. Using a combination of human strategy and creativity in synchronization with the advantages of algorithms could, therefore, present a plethora of new tactics and problem-solving for drug discovery, which has the potential to increase productivity.

4)Disease prediction

Experts use machine learning in health care primarily to determine which patients are at risk of developing certain conditions, like diabetes, asthma, heart disease, and other lifetime illnesses.

Figure-2.5 Disease prediction [5]

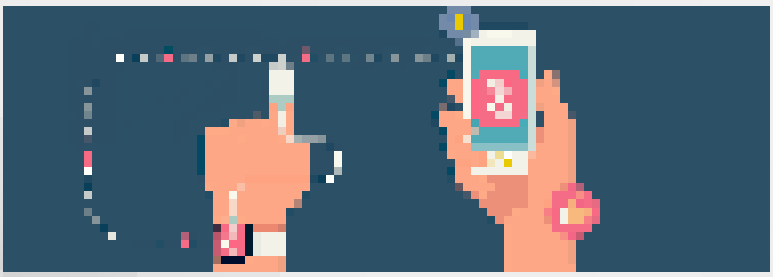


Figure-2.6 Monitoring Health [6]

5)Monitoring Health

The ability to monitor patients using AI may allow for the communication of information to physicians if possible disease activity may have occurred. A wearable device may allow for constant monitoring of a patient and also allow for the ability to notice changes that may be less distinguishable by humans.

6)Pattern detection and treatment suggestion in the spread of disease

Figure-2.7 Pattern detection [7]

There are numerous diseases, and AI has been used in a variety of ways to efficiently and accurately diagnose them. Some of the diseases that are the most notorious such as Diabetes, and Cardiovascular Disease which are both in the top ten for causes of death worldwide have been the basis behind a lot of the research/ testing to help get an accurate diagnosis. Due to such a high mortality rate being associated with these diseases there have been efforts to integrate various methods in helping get an accurate diagnosis'.

So, What makes Artificial Intelligence and Human Intelligence Different?

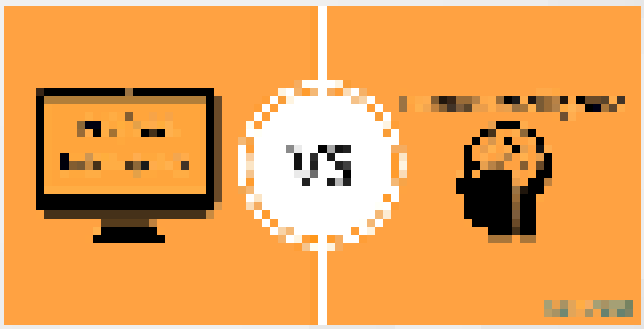


Figure-2.8 AI vs HI [8]

AI and Human Intelligence are different in many ways. AI has many advantages like the speed of execution of an AI system can not be matched by a human. AI provides the solution without any bias based on an opinion. AI can work with any kind of interruption due to fatigue. AI dominates any tasks that require monotonous execution. On the other hand, Human Intelligence excels at making a complex decision based on experience. Human Intelligence can manage multiple tasks at the same time, while AI has to learn a particular task every time. Human Intelligence also dominates the type of tasks that requires the use of abstract qualities and improvises much better than AI .

The role of AI in healthcare:-

Figure-2.9 AI in Healthcare [9]

Artificial intelligence is assisting doctors. According to Bloomberg Technology of Microsoft has developed AI to help doctors find the right treatments for cancer. There is a great amount of research and drugs developed relating to cancer. In detail, there are more than 800 medicines and vaccines to treat cancer. This negatively affects the doctors, because there are too many options to choose from, making it more difficult to choose the right drugs for the patients. Microsoft is working on a project to develop a machine called "Hanover". Its goal is to memorize all the papers necessary to cancer and help predict which combinations of drugs will be most effective for each patient. One project that is being worked on at the moment is fighting myeloid leukemia, fatal cancer where the treatment has not improved in decades. Another study was reported to have found that artificial intelligence was as good as trained doctors in identifying skin cancers. Another study is using artificial intelligence to try to monitor multiple high-risk patients, and this is done by asking each patient numerous questions based on data acquired from live doctor to patient interactions. One study was done with transfer learning, the machine performed a diagnosis similar to a well-trained ophthalmologist and could generate a decision within 30 seconds on whether or not the patient should be referred for treatment, with more than 95% accuracy.

Figure-2.10 Role of AI in Healthcare [10]

Can AI replace DOCTOR ?

The short answer is no. Although the field of Artificial Intelligence has made many advancements in the past couple of decades the systems currently in use are not very good when it comes down to empathy or reasoning. You train them to do a job and it will do it over and over like clockwork while getting more efficient at it. They can be excellent tools for doctors and nurses to deliver the data faster but the interpretation of the data is done by the doctor or nurse. Also, robots can not perform various complex task and surgery like a human. AI's job is to assist the professional rather than replacing them that has been the motive of AI since the start.

"Machines will do what human beings are incapable of doing. Machines will partner and cooperate with humans, rather than become mankind's biggest enemy."

- Jack Ma

To conclude I would like to say that AI will replace many jobs which require monotonous implementation due to which the efficiency of the whole organization is affected. But the jobs that require complex implementations which rely on abstract qualities and reasoning abilities will be benefited by the assistance of the AI.

Image Source:-

[1] https://miro.medium.com/max/1400/1*9xA-WzohT7d62LZHAD39Tg.jpeg
[2] https://miro.medium.com/max/604/1*csWRyhAA01dCj5YJlwxZkQ.jpeg
[3] https://miro.medium.com/max/474/1*JKTGlSDz7nyDZrujLQM6Hw.jpeg
[4] https://miro.medium.com/max/450/1*Yws8nn70kJmw0i2LHwg8Lw.jpeg
[5] https://miro.medium.com/max/450/1*V4SAUHlnTK6OBe-VcTC7Yw.jpeg
[6] https://miro.medium.com/max/575/1*-ffhUjD-EzMFR9sgFWIh6Q.png
[7] https://miro.medium.com/max/413/1*zzJCTcsHeNNcEITsyYrspQ.jpeg
[8] https://miro.medium.com/max/636/1*GdyiHxLijrp31NR5qOgmmg.png
[9] https://miro.medium.com/max/447/1*PwHKk5QyI9aYCKm_N3M-GQ.jpeg
[10] https://d2h0cx97tjks2p.cloudfront.net/blogs/wp-content/uploads/sites/2/2020/05/Role-of-AI-in-Healthcare-DF.jpg

3.Gradient descent

- Suraj Parmar
(ALUMNI, CSE, R.N.G.P.I.T)

3.1 Introduction:-

Gradient descent is used to find local minima of a given function. So, **It is a convex function based optimization algorithm.It is simply used to find the values of the parameters at which the given function reaches its nearest minimum cost.**

Intution:-

"A gradient is the ratio which relates the input and output of a function. How small changes drives changes in the output of the function." Suppose we have a function f(x) = x^2. Then the derivative of the function is 2*x. It means if the x is changed 1 unit then f(x) will change 2*1

Maths Proving This:

Let's take an example of Linear regression technique in machine learning, We have to find optimal 'w' and 'b' for the cost function J(w, b) where J is minimum.

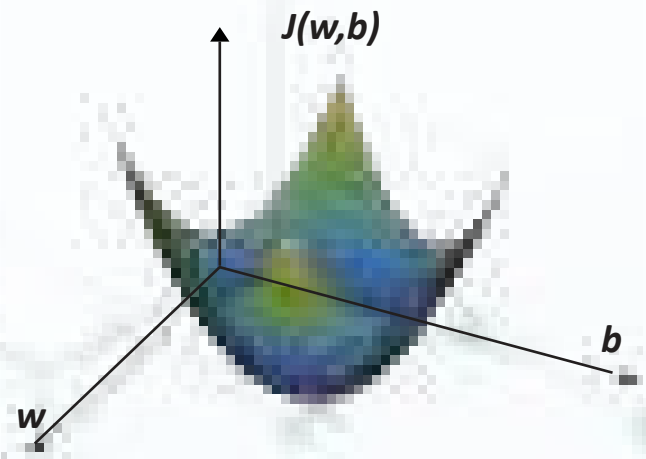


Figure-3.1.1 Linear Regression[1]

We start by taking random values for w and b somewhere near the top and then the gradient descent takes small steps towards steepest direction down till it reaches the point where the cost function is as small as possible.

3.2 Learning Rate:-

Figure-3.2.1 Learning Rate[2]

The steps which are taken to reach the optimal point decides the rate of gradient descent. It is often referred to as 'Learning rate'. To reach the local minima efficiently, we have to set them appropriately, neither too high nor too low. This is because if the steps(learning rate) are too big, it bounces between the convex function and may not reach the local minimum. If the learning rate is too small then gradient descent will eventually reach the local minimum but it will take too much time for that.

Figure-3.1.3 Comparison of Learning Rate[3]

To make sure Gradient Descent is working properly, we can plot(number of iterations vs. Cost function) the cost function as Gradient Descent iterates. This allows us to easily check how good our learning rate is. Just try and put all together. The cost function should decrease over time if the Gradient Descent is working properly.When Gradient Descent can't reduce the cost function anymore and remains near the same level, we can say it has converged. The number of iterations for convergence may vary a lot.

3.3 Types of Gradient Descents:-

PARAMETERS	BATCH GD ALGORITHM	MINI BATCH ALGORITHM	STOCHASTIC BATCH ALGORITHM
ACCURACY	HIGH	MODERATE	LOW
TIME CONSUMING	MORE	MODERATE	LESS

Figure-3.3.1 Types of Gradient Descent

Concave Vs Covex:-

It might happen that the cost function may not be non-convex. In this case we might end up on a local minima instead of a global minima.

Figure-3.3.2 Convex vs Concave [4]

Image Sources:-

[1] https://miro.medium.com/max/640/0*w67wX7xp5jWUISZy.png
[2] https://miro.medium.com/max/640/0*uykP7FsdR8P9MG6j.png
[3] https://miro.medium.com/max/640/0*IhGGqYM0KWDNpbb7.png
[4] https://miro.medium.com/max/320/0*L4A1EpH8AVhvARGk.png

4.Regression Using Sklearn

- Suraj Parmar
(ALUMNI CSE, R.N.G.P.I.T)

4.1 Introduction:-

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

4.2 Linear Regression:-

It is one of the simplest Algorithmto get started. The goal here is to fit a straight line between two or more variables. Where one is independent and another is dependent (i.e, Y=f(X), means that value of Y changes according to X but we can take any X in the range given.).

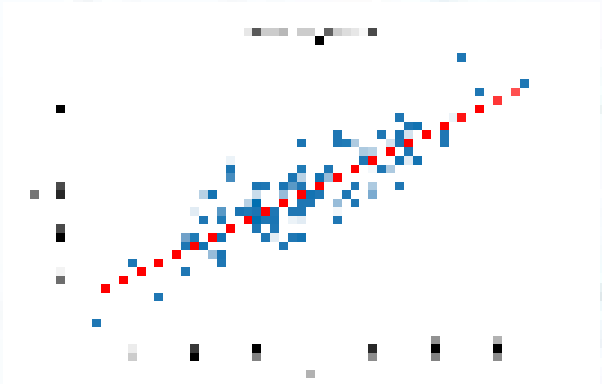


Figure-4.2.1 Linear Regression [1]

We try to predict real values at a given point. This gives us continuous values on a line

Thus, we try to find a relationship (definition of f(X)) between Y and X. Which is in the form of a line:-

Y = m*X + b.

Y — Dependent

X — independent

m — Slope

b — Intercept

Code:-

Figure-4.2.2 Code for Linear Regression

Multivariate Linear Regression:

It is to be noted that m can be multi dimensional, means X is multi-variate (we can represent in a table form instead of a 1D array). This is called 'Multivariate Linear regression'. We fit the line by reducing the Mean Squared Error (Loss) between predicted value and original Y. This is called Gradient Descent optimization Technique.

4.3 Logistic Regression:-

It is a classification algorithm though having 'regression' in its name. It is used to estimate discrete values (0/1, yes/no, true/false) from the independent variables. It predicts the probability of an event by using a logistic function. so its also known as logit regression. We know that probabilities lies between 0 and 1, so it predicts value between 0 and 1(like regression).

$\text{logit}(p) = \ln(p/(1-p))$

Code:-

Figure-4.3.1 Code for Logistic Regression

Image Sources:-

[1] https://miro.medium.com/max/400/0*qhW46ylLs0BK_xmm.png

5. Machine Learning in the Browser using TensorFlow.js

- Suraj Parmar
(ALUMNI, CSE, R.N.G.P.I.T)

5.1 Introduction:-

Figure-5.1.1 TensorFlow.js Library

TensorFlow.js is a library for developing and training ML models in JavaScript, and deploying in browser or on Node.js and can utilize WebGL to run on an underlying GPU.And also all data stays in users' device and no need to send data to servers.

What can you do with TensorFlow.js?

If you're developing with TensorFlow.js, here are three workflows you can consider.

- You can import an existing, pre-trained model for inference.
- You can re-train an imported model.
- Author models directly in browser.

Lets Code:-

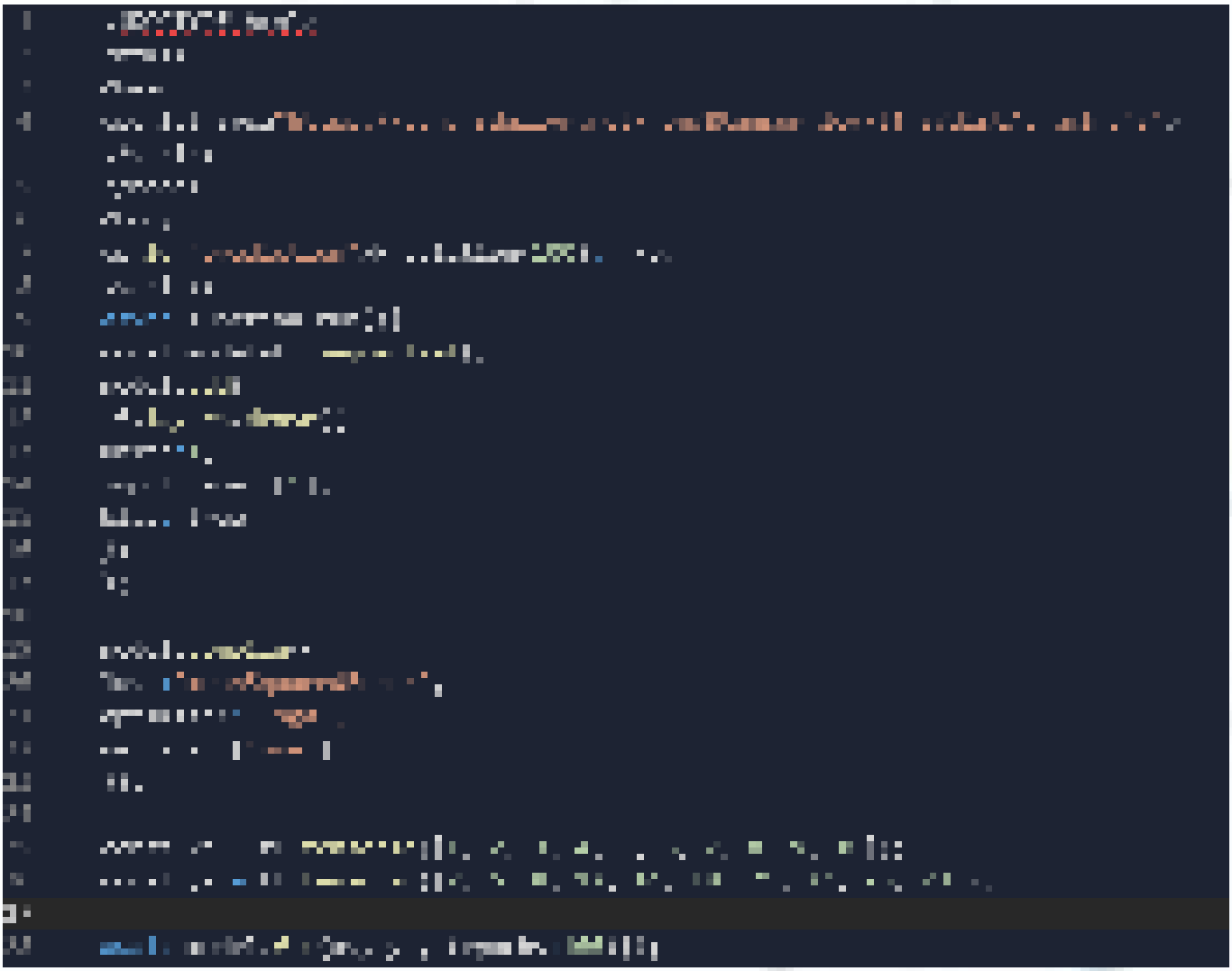


Figure-5.1.2 Code for Basic Webpage

5.2 Layer and Compiling:-

Layer:- Here, we have added a Dense layer(Fully connected) with only single neuron,which takes only single number(like Scaler) for prediction and perform inference on it outputting a number which should be close to 3*(input)-1.

Compiling:- The API supports all Keras layers (including Dense, CNN, LSTM, and so on). Then we have to say the model to use which Optimization technique and a Loss function. We can give metric as a parameter if we want.

6. Beginners with Data Analytics in Machine Learning..?

- Shreya Lad
(FOURTH year, CSE, R.N.G.P.I.T)

6.1 Introduction:-

Technical definition of ML - "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience."

"The only source of knowledge is Experience."
-Albert Einstein

Figure-6.1.1 Data Analytics

Machine learning is the subset of Artificial Intelligence(AI).
Deep Learning is the subset of Machine Learning.

What is Artificial Intelligence (AI)?

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers “smart”. They work on their own without being encoded with commands.

6.2 Machine Learning in Data Analytics:-

Machine learning is a method of data analysis that automates analytical model building. Analytics involves studying historical data to research potential trends , to analyze the effects of certain decision or event or to evaluate the performance of a given tools or scenario. The goal of analytics is to improve the processes by gaining knowledge which can be used to make improvement or changes.

Figure-6.2.1 Graph of Difficulty vs Value in Machine Learning Data Analytics [1]

What is the path of Evolution of Machine Learning?

Figure-6.2.2 Path of Evolution of Machine Learning

6.3 Types of ML Algorithm:-

Figure-6.3.1 Types of ML Algorithm [2]

Supervised Learning:- In supervised learning, ML algorithm is given a small training dataset to work with. This training dataset is a smaller part of the bigger dataset and serves to give the algorithm a basic idea of the problem, solution, and data points to be dealt with. The training dataset is also very similar to the final dataset in its characteristics and provides the algorithm with the labeled parameters required for the problem.

Unsupervised Learning:- It holds the advantage of being able to work with unlabeled data. This means that human labor is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program.

Reinforcement Learning:- It directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial and error method. Favorable outputs are encouraged or 'reinforced', and non-favorable outputs are discouraged or 'punished'.

Image Sources:-

[1] https://www.governanceanalytics.org/knowledge-base/images/image001.png
[2] https://miro.medium.com/max/2000/1*8wU0hfUY3UK_D8Y7tbIyFQ.png

7. AutoEncoders-Denoising Understanding

- Suraj Parmar
(ALUMNI, CSE, R.N.G.P.I.T)

7.1 Introduction:-

"An autoencoder is a neural network that is trained to attempt to copy its input to its output."

-Deep Learning Book

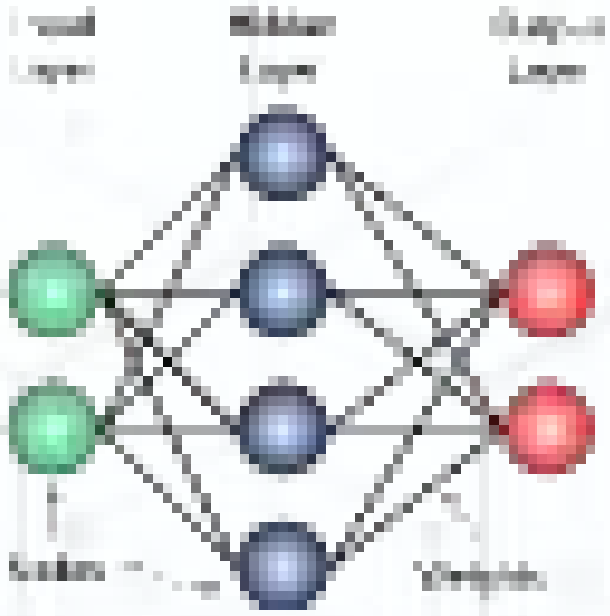


Figure-7.1.1 Layers of Encoders

Here we can see that we are trying to first learning to map input to a bottleneck then the bottleneck to output. This is an end to end process.

Encoder: Takes an input x (this can be image, word embeddings or voice data) and produces an output h.
For instance, Think of an image of dimensions 32x32x-1(HxWxC) and it is reduces to a 3x1output.

Decoder: Takes an input h(densly represented) and produces an output ~x. For example, 3x1 vector as input produces a 32x32x1 image which resembles x.

The ideal autoencoder model balances the following:

- Sensitive to the inputs enough to accurately build a reconstruction.
- Insensitive enough to the inputs that the model doesn't simply memorize or overfit the training data.

7.2 Visualization of Features and reconstruction:-

Figure-7.2.1 Visualization of Features [1]

Better results by using **Convolutional Neural Nets** CNNs have proven to be very good with image data. So instead of flattening images, We can directly feed images to CNN encoders and decoders.We can expect to see better results by using them.

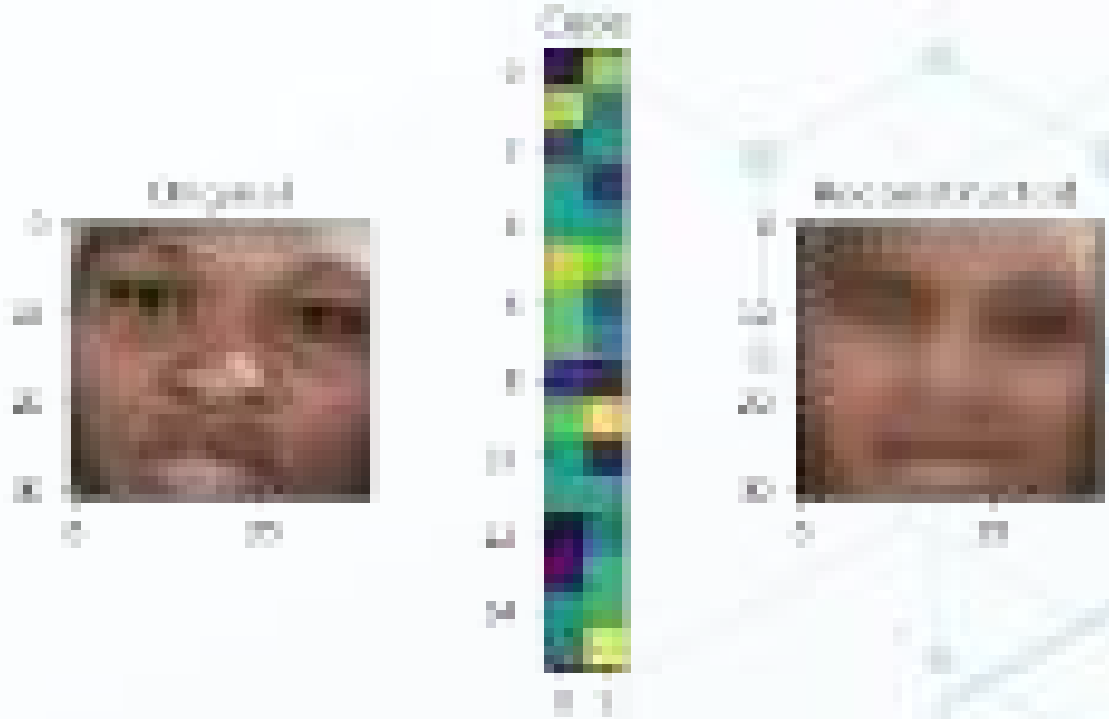


Figure-7.2.2 Reconstruction [2]

The idea here is that ConvNets learn better by learning a better way to represent input into a latent space by learning more features using filters.

Figure-7.2.3 Reconstruction [3]

We can clearly see that the reconstruction results are better compared to previous deep AE. And we can see the latent space representations which is a compressed representation of the input. Every image passes from same encoder function and is compressed. Which leads to smaller size image. We can perform clustering with less compute on it.

A fun application — image denoising:-

Since AE can learn to represnt images into a latent space and reconstruct from it. It can also learn to remove noise from images. For example,we train it to map noisy images to clear images.

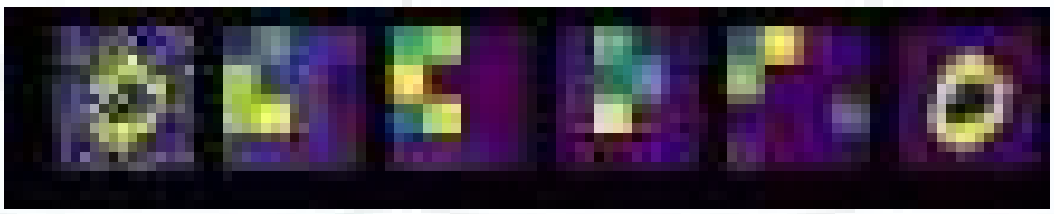


Figure-7.2.4 Noisy image reconstruction [4]

It seems to work pretty well. If you scale this process to a bigger ConvNet, you can start building document denoising or audio denoising models.

Sequence to Sequence:-

If your inputs are sequences, rather than vectors or 2D images, then you may want to use as encoder and decoder, a type of model that can capture temporal structure, such as a LSTM. To build a LSTM-based autoencoder, first use a LSTM encoder to turn your input sequences into a single vector that contains information about the entire sequence, then repeat this vector n times (where n is the number of timesteps in the output sequence), and run a LSTM decoder to turn this constant sequence into the target sequence.

Variational AutoEncoders:-

Instead of letting Neural net decide how to represent in latent space, We put constraints on that. Means, limiting latent space representations so we can reconstruct using very few parameters. We will explore that later.

Image Sources:-

[1] https://blog.keras.io/img/ae/autoencoder_schema.jpg
[2] https://stackabuse.s3.amazonaws.com/media/image-reconstruction-and-denoising-with-autoencoders-in-python-and-keras-7.png
[3] https://miro.medium.com/max/893/1*YZpPi3xDzwD00trn3r0Fcw.png
[4] https://miro.medium.com/max/893/1*yZ2C59aKXARCA7qVnKflYQ.png

8. Significance of p-value in Machine Learning

- Visarg Desai
(ALUMNI, CSE, R.N.G.P.I.T)

8.1 Introduction:-

Figure-8.1.1 Data Quote[1]

It is a very pleasant day. With the aim of learning something new in your machine learning journey, you start learning multiple linear regression. But before you even start, you come across this mystery called the p-value. This term will make you go through high school mathematics and even after a significant amount of time, you will not understand its meaning. But it is a very important concept involved in training your model. Let's understand this mystery through the famous coin toss example.

One day a friend comes to your house to show you a magic trick. He brings a coin with him. Before the magic trick starts you think that the coin is fair (both head and tail are present). This becomes the null hypothesis. Here we are trying to prove that the null hypothesis is wrong. And our alternative hypothesis is that the coin is not fair (2 heads or 2 tails).
Now your friend starts the magic trick.

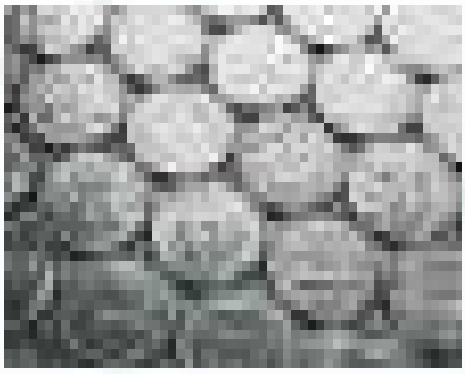


Figure-8.1.2 Coins [2]

8.2 Examples:-

Figure-8.2.1 Output

First Toss:-

He tosses the coin first time and it lands on Head. This is not a surprising outcome as the probability of this happening if the null hypothesis is true is 50 % (0.5).

Second Toss:-

He tosses the coin second time and it lands on the head again. There is still a fair possibility (25 % or 0.25) of this happening. Till now, our null hypothesis is true and we have no doubt.

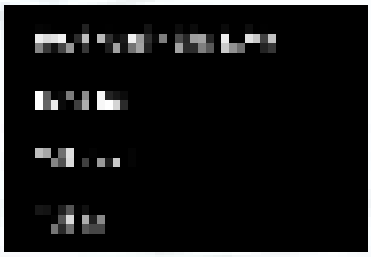


Figure-8.2.2 Output

Figure-8.2.3 Output

Third Toss:-

Now for the third time, it lands on the head. Your confidence in the null hypothesis starts decreasing because the possibility of 3 heads in a row is very low (~12 % or 0.12) if the coin is fair. You give benefit of doubt to him as he is your friend.

The another toss can make you doubt your friend. This time too, the coin lands on the head. You start doubting that the coin might be a magic shop coin because the possibility of this happening is low (~6% or 0.0625,~3% or 0.03125) and you now reject the null hypothesis that the coin is fair because of low signification. You decide to go with alternative hypothesis of the coin being unbalanced. And it turns out that the coin was bought from a magic shop.

Is p-value the probability that the null hypothesis is true?

No, Conclusion that the null hypothesis is true. It is also possible that the null hypothesis might be true even if the p-value is less. It depends on the data you are considering

What is the significance of p-value in machine learning?

Let's apply the above theory to a dataset. Consider a dataset with different independent variables (features) such as No. of Hours, Attendance, and Gender of the student (0 for male and 1 for female). Based on these 3 features, we have to predict the output, i.e., the marks obtained by the student.

The equation for multiple linear regression looks like this:-

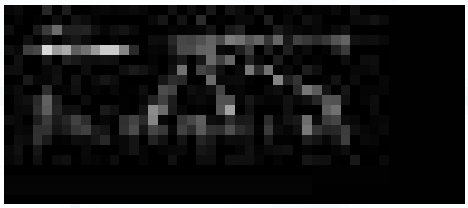


Figure-8.2.4 Output

Here, our **null hypothesis is that there is no relation between the features and output.** This means that there will be no effect on the output if the values of features are changed. And then there is our **alternative hypothesis.**

P-value decides the value of the coefficient present in the multiple linear regression equation. the coefficients will be 0(due to no relation between features and output). The p-value will decide whether this is true or not.

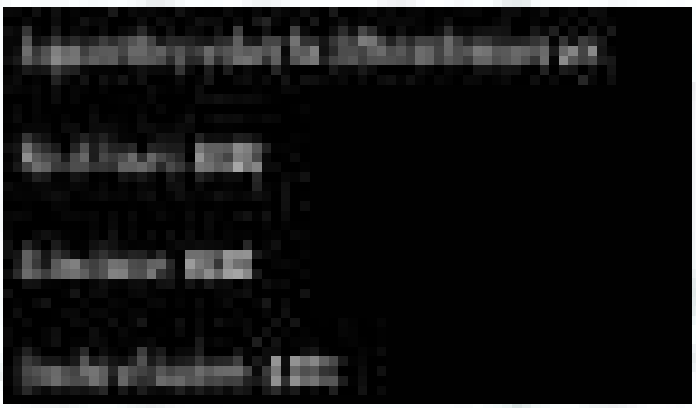


Figure-8.2.5 Output

But before that, we need to set the significance level of the p-value. Let's go with the usual value p=0.05, This value says that any feature with a p-value less than 0.05 will prove the null P-values hypothesis to be wrong.

The above values denote that hours and attendance have a p-value lower than the significant level (0.05). But gender has a p-value greater than the significant level (0.05). This proves that our null hypothesis is wrong and that there exists a relationship between the features (hours and attendance) and the output (Total marks). It also suggests that the gender of the student follows the null hypothesis (there is no relationship between the gender of students and marks scored by them (which is true). This means that changes in **hours and attendance (Features)** will trigger a significant change in **total marks (Output).**

Image Sources:-

[1]https://miro.medium.com/max/1050/0*oK22FtpIqvCxGpp6
[2] https://miro.medium.com/max/1050/0*Li6C3C7AbbVdRBWZ

9. Data Science

- Visarg Desai
(ALUMNI, CSE, R.N.G.P.I.T)

Data scientists and data analysts are well-known to the general public. However, there is one role that many of us confuse with the other two. And the title of the position is Data Engineer.

PYTHON

A Data Analyst analyses data and generates a report that details the strategies that can benefit the company. Data scientists work with massive amounts of data and apply their knowledge of arithmetic, statistics, programming, and machine learning to solve problems.

Data is acquired from a variety of sources in the actual world. This information cannot be utilised to train models directly. First, some processing is required to structure and uniformize the data set.

Figure-9.1 Data Engineer
[https://miro.medium.com/max/960/1*Dmk8Vi4G2r0KNfWYcVyb9Q.png]

By obtaining data, a Data Engineer creates a vast pool of data and ensures a smooth data pipeline. They work on a variety of tasks, including cleaning up filthy data, creating data pipelines, and data

Consider this data set. Lets perform some Data Science!!

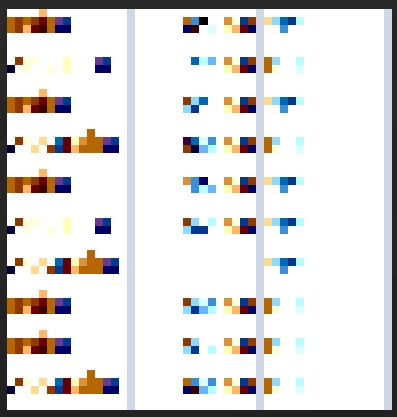


Figure-9.2 Data Set

Step 1:- Importing the python libraries:-

Figure-9.3 Code of Importing Python Libraries

Pandas is one of the most widely used python library for data science. It provides easy to use structures and data analysis tools. Its powerful object known as pandas data frame are very much helpful in data analysis. NumPy is a Python library which provides support for large multi dimensional array and allows to perform mathematical operations on them.Apart from these 2, matplotlib is also a much used library for visualizing the data in form of graphs, histograms, etc.

Step 2:- Importing the data set and structuring using Pandas:-

After importing libraries, we need to get the datasets on which further work will be done. The Pandas library is used to import the dataset. Using read_csv() function, the data set gets converted to pandas data frame. Pandas data frame is a 2 D table object.

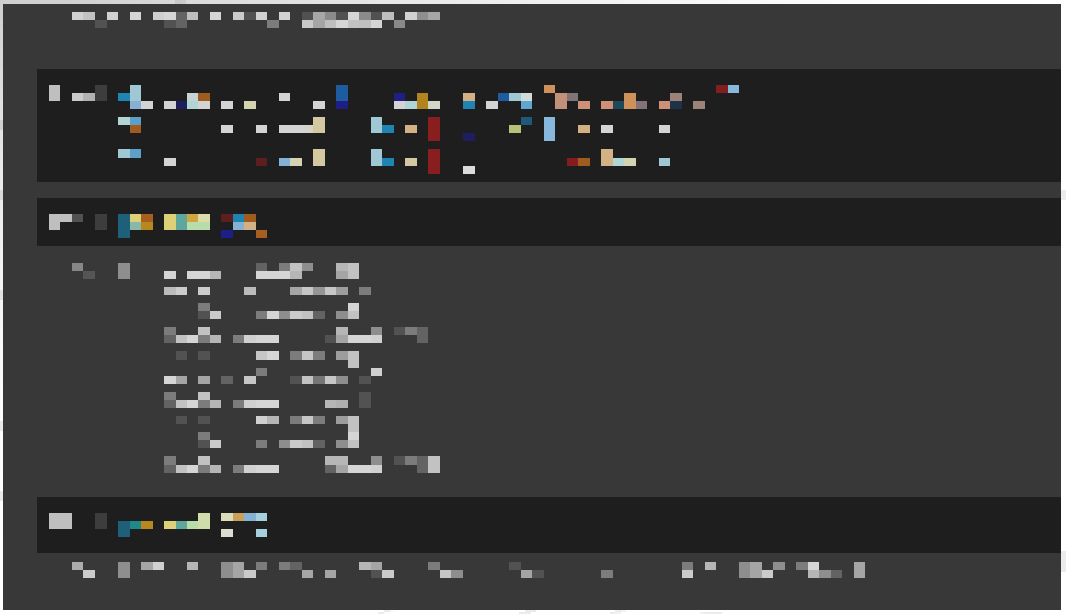


Figure-9.4 Code of Importing DataSet

This data set can be divided in two parts :-

i) Features: Features are the independent properties of the data which have effect on the output. Here, Country and Salary are the features.

ii) Output. Output can be considered as a dependent variable which depends on features.Here,Answer is Output

Here, Country and Salary are the features. And Yes or No are the output dependent on these features.

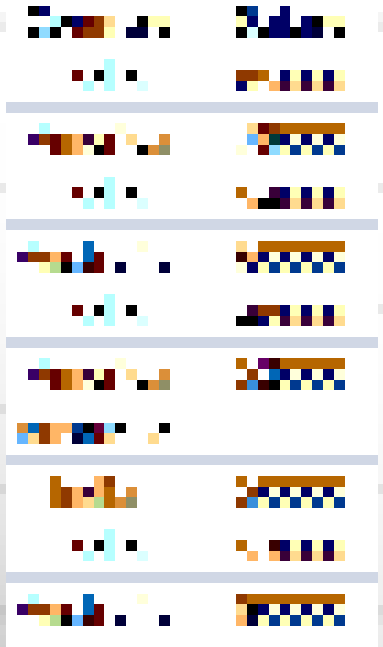


Figure-9.5 Features

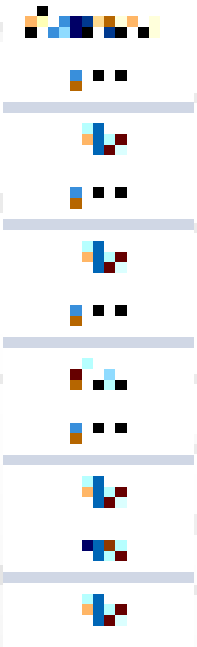


Figure-9.6 Output

Once the pandas data frame is generated, iloc is used to specify rows (As seen in the code snippet [:,:-1] means all the rows from start to end and all the columns except the last) and columns on the basis of its integer index. It is used to separate features and Output in our data set. Here X contains all the features. And Y contains the dependent variable.

There are few problems in this raw data. It need to be cleaned before we use it for analysis. The issues are:-

1. Columns should be renamed to a more readable name.
2. Country name with multiple words consists of underscore instead of white space.
3. Some country names have brackets in the end which are not useful.

Renaming Columns:-

One way to rename the columns is to use the **rename()** function and pass a dictionary to it with **old names as key and new names as value.**

Figure-9.7 Code of Renaming Columns

Another way is to just initialize the column names without caring about the old name. We can do it by creating a list with new names written in the same sequence as columns.

Figure-9.8 Code of Renaming Columns

Replacing underscores in the country names:-

Apart from this image there are many other country names with underscore instead of white spaces. To fix this, we can use replace() function on the Country column.
The code to perform this task can be written as:

Figure-9.9 Code for Replacing Underscore

Here str is used to convert the column value into string so that we can then replace the sub-string using the replace function.

Removing brackets:-

As seen in the image, some country names consists of brackets. These brackets will not give us any advantage while analyzing data, so we will remove this whole bracket using regex.

Figure-9.10 Code for Removing Brack-

Here the regex '\(.*\)' denotes everything in the bracket including the brackets. We will replace this regular expression with empty string using the replace function. The rstrip() function is used to remove any trailing spaces.

Step 3:- Replacing missing values:-

In real world, it may happen that some feature values are missing which may create problems in future. So, we need to take care of that. It can be handled in two ways :-
i) Delete the entire row or
ii) Replace the missing value.
First option may result in loss of data so it is not preferred. So lets stick with the latter one.

SciKit(sklearn) library contains lots of efficient tools for ML and statistical modeling. **Impute** package of sklearn provides a class named **SimpleImputer** which can be used to replace the missing values with mean, median or most frequent value of column.

Figure-9.11 Scikit Learn

Figure-9.12 Code for Importing SimpleImputer

Here we have used mean strategy which indicates that all the missing values in a column will be replaced with mean of that column. The **missing_values** parameter equal to np.nan represents missing data which needs to be replaced. After we select a strategy to replace missing values, **fit()** method is used to calculate values for all the missing entry while **transform()** method is used to apply these values wherever the entry is missing. This can also be achieved in one step using **fit_transform()** method.

Step 4:- Encoding features and output:-

The problem with this data is that the machine cannot understand country names as features. So there is a need to encode them into numeric values. The preprocessing package of sklearn provides **LabelEncoder** and **OneHotEncoder.**

LabelEncoder transforms the categories present in features into successive numeric values. In our case, India-0, America-1, Australia-2. But due to this, machine starts making comparison by thinking of this as an order (India < America < Australia). It may happen that the machine will think that average of India and Australia is America. So this approach should not be used for such features.

Figure-9.13 Code for Encoding Features

What OneHotEncoder does is it creates n different columns for n different categories in a feature. So, in our case 3 different columns are generated, one each for America, Australia and India. Now, in the raw data for all the rows having India as country, the column created for India after OneHotEncoding contains value 1 and the other 2 columns have value 0.

ColumnTransformer transforms the specified column using OneHotEncoder and concatenate the result with remaining columns. Here, remainder='passthrough' indicates that all other feature columns remains unchanged.

Figure-9.14 Code for ColumnTransformer

Now, using fit_transform we calculate and apply the numeric values on the features and convert them into a NumPy array. In our case, the output has only 2 values , i.e. , Yes or No. So, Label Encoder can be used to encode them into 0 and 1.

Step 5:- Feature Scaling:-

It may happen that the range of each feature is very large. It is observed that there is a vast difference between minimum value and maximum value of the feature. Such features needs to be scaled so that all the feature fall under the same range. It can be done in two ways:Normalization and Standardization. Both methods has their pros and cons. So, it depends on what suits best for your model.

Figure-9.15 Code for Feature Scaling

The preprocessing package of sklearn provide StandardScaler class which performs standardization. And then as always fit_transform() is used to calculate and apply the scaled values.

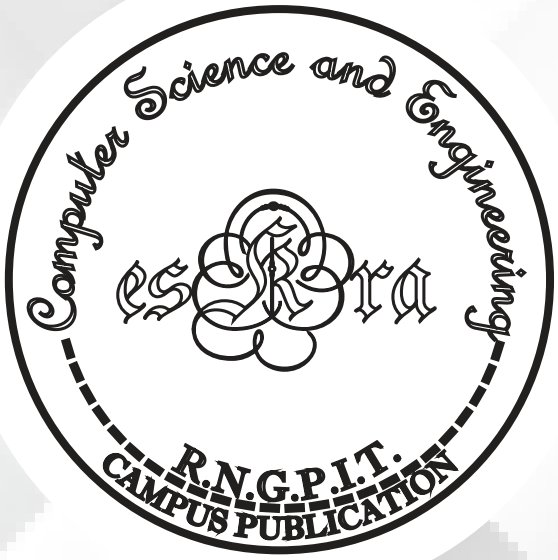
Step 6:- Splitting data into training and test sets:-

Now comes the final step of splitting the data set into 2 parts. Over-fitting may occur if there is a lot of data to be trained on which results in a complex hypothesis (predictor). Under-fitting may occur if there is less data available for training the model. So, a healthy ratio should be maintained between training and test. .data.

Figure-9.16 Code for Splitting Data

10. Roadmap

Student Co-ordinators



www.eskra.rngpit.ac.in

A Magazine by:
ESKRA - Campus publication
Computer Science and Engineering,
R.N.G. Patel Institute of Technology,
Surat

You can provide feedback on
eskra.cse@rngpit.ac.in

eskra.cse@rngpit.ac.in

cserngpit

cserngpit

+91-9427527398
+91-9033707334
+91-9510427118