Machine Learning with AI
By Ankit Kumar

Basic Skills:
python
numby basics
pandas basics - csv data ko manuplate karna
additional - tensor flow, seaborn, sicpy,
statistics & probability - alzebra
ml basics - linear regression, gradient descent
supervised k saare algo
aur unsupervised k saare algo
clustering ko implent karna scikit learning me
scikit learning, tensor flow

Data
Data is collection of raw facts and figures
Data past -> small, structured
Data now -> huge, unstructured

Data science lifecycle
ek process hai jo data ikattha karne, saaf karne, analyze karne, model banane, usko test karne aur use karne tak jaata hai.

Machine learning
jismein computer ko data dekar cheezein sikhai jaati hain, taaki woh apne aap seekh sake

Ml Pros
manual effort ke bina accurate predictions aur decisions leta h.
future predictions me, pattern ko recognise karne me, human ko efficiency ko kam karne k liye, automation me, continuous improvement, handling lot of data, wide apps

Ml Cons
yeh galat predictions kar sakta hai, large amount of data ki jaroorat padti h, bhot saare resouse ki jaroorat padti h, bhot saare time ki jaroorat padti h, data ko collect aur dhundna padta h, aur unpar kaam karna padta h, data ko process aur filter karne par time consume hota h

Ml Uses
Ml ka use cheezon ko samajhne aur unke bare mein bhavishya mein anuman lagane ke liye hota hai.
language translation, medical diagnosis, stock market, online fraud detection, virtual personal assistant, email spam & malware filtering, tesla car, traffic prediction, speech recognition, image recognition

Ml Facts
aisa software hota hai jo jisme woh khud hi data se seekh kar kuch naya kar sakta h

PATTERN
jo computer data dekh kar samajh leta hai aur phir naye data par use karta hai.
data ke regularity ya structure ko describe karta hai jo machine ko future predictions karne mein help karta hai

Features
features wo data points hote hain jo humare model ko patterns samajhne mein madad karte hain
features wo characteristics hote hain jo humare data mein maujood information ko describe karte hain aur jo humare model ko training ke liye use kiye jaate hain
ml me attributes aur class features hote h

Labels
वो निशान होते हैं jinhe data par lagate hain जो डेटा को परिभाषित करते हैं,een labels se जिससे मॉडल सीखता है।

Data collecition
Machine ya model ko sikhane ke liye zaroori data ikattha karte hain.

Hame ml me data kha se milega
1. company se, 2. investor se, 3. online khud data ko dhund sakte h, 4. uci ml

UCI ML
UCI stands for the University of California Irvine machine learning repository, and it is a very useful resource for getting open source and free datasets for machine learning.
jisme ye new or old data provide kiya jaata h

UCL
UCL (Upper Confidence Limit) woh value hoti hai jo batati hai ki ek prediction ki upper limit kya ho sakti hai, yani prediction kitni high ja sakti hai.

Typical machine learning problems
Typical machine learning problems mein classification, regression, clustering, aur recommendation tasks hoti hain
Machine learning problems mein, cheezein categorize karna, values predict karna, similar cheezein group karna, aur suggestions dena shamil hote hain.

Ml k types
Supervised learning
Supervised learning machines को सुधारे हुए उदाहरणों से सीखने और भविष्यवाणी करने में मदद करता है ex. purane gharo ki price ko dekhkar naye gharo ki price predict karta h
Supervised types - regression, classification
sl future prediction k liye, recommended system banane k liye istemal kiya jata h
Jisme ye khud teacher hota h aur decide karta h
Ise reward aur feedback milta h
already tagged data
features and labels present
classification

regression - value dene wala h

Unsupervised learning
mix data k pattern aur nature ko dekhkar alag alag groups banata h
algorithm ko data ke patterns khud se discover karne padte hain bina kisi labelled guidance ke.
ek fruits basket me alag alag fruits ko category k basis par divide kar dete h
isme group ka kaam hota h
clustering
association

REINFORCEMENT LEARNING
ml hamare next movement par decision leta h ek ai dimag hota h jo human se sikhta h
ex. jese human vs computer ludo game
day to day life me use kiya jata h
Reware based
Feedback based

Classification task
ek model ko data diya jata h jisse oos data k aadhar par naye data ko kisi class mein classify kiya jata hai
data ko different categories mein classify kiya jata h
Spam filter: email ko spam ya not spam mein classify kiya jata h
Image recognition: ek image ko cat ya dog mein classify karna

Regression
ek supervised learning technique hai jo continuous values ki prediction karne ke liye data points ke beech linear relationship model karta hai.
jo data mein patterns ko dhoondhta hai aur un patterns ka use karke continuous values ki prediction karta hai.
$Y = f(x)$

Regression task
ek continuous outcome (jaise house prices) ko predict karne ke liye ek ya adhik predictor variables (jaise square footage) ka use karte hain.
Stock ki kimat ki bhavishyvani karna
Ghar ki kimat ki bhavishyvani karna
mausam ki bhavishyvani karna

Batch learning mein
model ko ek baar mein poora dataset diya jaata hai, aur fir model apne parameters ko optimize karta hai.
gradient descent algorithm ka istemal kiya jaata hai.
model ko ek baar mein sabhi data dekar optimize kiya jaata hai.

Online learning technique
Model ko train karne k liye model ko thoda-thoda data feed karke sikhaya jaata hai.

Root Mean Square Error
ek aisa metric hai jo humein yah batata hai ki kisi model ke predictions kitne accurate hain. RMSE yeh bata deta hai ki model ki predictions aur actual values ke beech mein kitna average difference hai.

scikit learn
Scikit-learn ek Python library hai jo machine learning algorithms ka ek bada set pradan karti hai, jaise ki regression, classification, clustering, dimensionality reduction, data pre-processing, etc.
Scikit-learn is a Python library for machine learning.

***using sklearn for ml**
simple and effiecient tools for data mining and data analysis
accessible to everbody, and reuse in various contexts
built on numpy, scipy, and matplotlib
open source , commericially,
usable - bsd license

training
ka matlab hai ek computer ko data se sikhana taki woh apne aap tasks kar sake aur apni accuracy badha sake.**

test splitting
data ko do भागों में बांटना है: एक मॉडल को सीखने के लिए और दूसरा यह जांचने के लिए कि मॉडल नए डेटा पर कैसा काम करता है।
Machine learning mein test splitting ek aisi prakriya hai jismein data ko do hisson mein baant diya jata hai - ek training set aur ek testing set.
Training set, model ko sikhaane ke liye istemal kiya jata hai.
Testing set, model ke performance ka mulyankan karne ke liye istemal kiya jata hai.

house prices
Machine learning mein house prices, gharon ki keemat ka anuman lagaane ke liye, alag-alag karkon ka use karke ek statistical model taiyar karna hai.
Yeh model, ghar ki location, size, features, aur market ke conditions jaise data ko use karke, ghar ki sambhavit keemat ka anuman laga sakta hai.

Linear regression
ek machine learning model h jo straight line use karta h
jisse value ko predict kiya jata h jo dusre value par based hoti h
kese kaam karta h?
Linear regression ek supervised learning algorithm hai jo do variables ke beech linear relationship ko model karta hai aur unke beech ek line fit karta hai jo unke beech ke data points ke best fit ko represent karta hai.
Yeh model naya data ke liye target variable ki value predict karne ke liye use kiya ja sakta hai.
loss function linear regression me sum of squared error

Simple Linear Regression

Ek independent variable par Sare points per line fit karta h
ek statistical method hai jo ek d ependent variable aur ek independent variable ke beech linear relationship ko model karta hai.
yah ek straight line ka use karta hai jo do variables ke beech ka sambandh dikhaata hai.
Simple Linear Regression ek aisa algorithm hai jo ek dependent variable ko ek independent variable se jodne wali ek seedhi rekha ko dhoondhta hai.
Advantages
simple to understand
simple mathmatical model
explain to someone easily
Disadvantages
Y = a + bx

Measure of good fit
Sigma (y0-yp)^2 = no.

Line eqn -> y = mx+c

Multiple Linear regression model
Ye plaine, approximately lagakar equation ko pura karta hai
ek aisa model hota hai jo ek dependent variable (निर्भर चर) ki prediction karne ke liye ek se adhik independent variables (स्वतंत्र चर) ka use karta hai.
Multiple regression model ek aisa statistical model hai jo ek dependent variable ke liye ek ya ek se adhik independent variables ke beech linear relationship ko model karta hai.
Y = ax1 + bx2 + cx3 + dx4 + e

loss function
Machine learning mein, loss function yeh batata hai ki aapki predictions kitni galat hain
loss function yeh batata hai ki model kitna galti kar raha hai apne predictions mein
loss function model aur actual data ke beech mein kitna difference hai yeh measure karta hai

gradient descent
Machine learning mein gradient descent kisi function ko kam karne ke liye bar-bar uski dhalan ki disha mein chhote kadam lene wala तरीका (tarika) hai.)
Gradient descent is an optimization algorithm commonly used in machine learning to find the minimum of a function by iteratively moving in the direction of steepest descent.

mini Batch Gradient Descent
mini Batch Gradient Descent ek optimization technique hai jo machine learning mein model parameters ko optimize karne ke liye training data ke chhote chhote batches ka use karta hai.
Mini-batch gradient descent machine learning mein aisa algoritham hai jo training data ke chote chote groups (mini-batches) istemaal karke model ke parameters ko update karta hai.

why use gd and loss functions
1. exact methods computationally expensive
2. loss functions give direction of optimal solution

3. fast enough to scale on big data
4. easy to understand

mean squared error
hamari value kitni satik h

import matplotlib.pyplot as plt
isse features aur labels ko plot karte h

plot karne ki shuruat karte h
plt.scatter(diabetes_x_test,diabetes_y_test)
plt.show()

Batch time
Machine learning में, batch time वो समय है जो एक डाटा के छोटे हिस्से को प्रोसेस करने में लगता है।
Machine learning mein batch time, ek training dataset ke ek subset ko process karne ke liye lagane wala samay hota hai

Yeh ek chhota hissa hota hai jo ki puri dataset se nikala jata hai aur model ko update karne ke liye use kiya jata hai.

Batch size, batch time ko nirdharit karta hai, aur yeh hyperparameter hai jo ki model ki training ko affect karta hai.

Gradient Descent
ek optimization algorithm hai jo machine learning aur deep learning models ko train karne ke liye cost function ko minimize karta hai.
Yeh algorithm model ke parameters ko iteratively adjust karta hai, yeh dekhte hue ki unme se kaun se cost function ko sabse zyada kam karte hain.
Yeh ek bahut hi important algorithm hai jo kai tarah ke machine learning tasks mein kaam aata hai.
Machine learning में, gradient descent एक ऐसा तरीका है जो बार-बार गणित गणना करके किसी मॉडल को बेहतर बनाने में मदद करता है।
Gradient Descent ek तरीका hai jaldi se kisi function ke sabse kam mulya ko dhundne ka, galtiyon ko kam karke model ko seekhne mein मदद ke liye.)

Mini-batch gradient desecent
is a training method for machine learning models that uses small batches of data to update the model's parameters, balancing speed and accuracy.
Mini-batch gradient descent machine learning models को train करने का एक तरीका है जिसमें data को छोटे groups में बाँटकर उन groups पर बार-बार updates दिए जाते हैं।
Mini-batch gradient descent machine learning models को train करने का एक तरीका है जिसमें data को छोटे-छोटे groups में बाँटकर थोड़ा-थोड़ा करके सिखाया जाता है।
model ko train karne ke liye gradient descent optimization technique ka use karta hai, jismein dataset ko chhote chhote batches me baantkar unka use karke model parameters ko update kiya jata hai.

Stochastic Gradient Descent
(SGD) ek optimization algorithm hai jo machine learning models ko train karne mein use kiya jaata hai, khaskar badi datasets ke liye. Yah traditional gradient descent se zyada efficient hai kyunki yeh har iteration mein complete dataset ki jagah sirf ek random batch ya single data point ka use karta hai.
Stochastic Gradient Descent (SGD) ek optimization algorithm hai jo machine learning models ko train karne ke liye use kiya jaata hai, yeh ek baar mein poora dataset ki jagah chhote chhote batches ya ek data point ka use karke gradient ki calculation karta hai aur model ke parameters ko adjust karta hai.
Yeh bada datasets ke liye bahut efficient hai aur neural networks ko train karne ke liye standard algorithm hai.
Yeh SGD ko badi datasets ke liye practical banata hai aur yeh deep learning jaise complex models ko train karne mein bhi madadgar hai.

classificaton of problem
me pata hota label kya honge, inme se konsa object h

regression of problem

Supervised learning classification
involves training a model using labeled data to predict the category of new, unseen data points
Supervised learning में classification डेटा को predefined categories में विभाजित करने के लिए एल्गोरिथम को訓練 (train) करना है।
Supervised learning ka vargikaran data ke lebel wale udaharanon se seekhkar nayi cheezon ko shreniyon mein vargikrit karna hai.

Supervised learning regression
ek technique hai jismein ek algorithm ko input aur output data ke jodiyon ke ek set ka upyog karke sikhaya jata hai, taki woh naye input data ke liye output ki bhavishyvani kar sake.
Yeh ek sentence mein regression ko samjhane ka ek tarika hai.
Agar aap aur janna chahte hain, to main aapko bata sakta hun ki:
Regression algorithm kaise kaam karte hain
Alag alag prakar ke regression algorithm hain
Regression ka upyog kaun si samasyaon ko hal karne ke liye kiya ja sakta hai
Supervised learning regression में, किसी उदाहरण को देखते हुए उसका संख्यात्मक मान (जैसे, घर की कीमत) आंकड़ों से सीखकर अनुमान लगाया जाता है।

K-Nearest Neighbors
(KNN) classification is a machine learning technique that predicts the class of a data point based on the majority vote of its closest neighbors in the training data.
ek supervised learning algorithm hai jo **similar data points** ko **k nearest neighbors
Yeh algorithm **non-parametric** hai aur **lazy learning** approach ka use karta hai.
ek machine learning algorithm hai jo **similar data points** ko ek hi group me classify karta hai, **distance** ke aadhar par.
Example: Agar aap **flower** ki classification karna chahte hain, to KNN algorithm **new flower** ko **sabse similar flowers** ke group me classify karega, jo training data me hain.

K Nearest Neighbors (KNN)
is a machine learning algorithm that classifies new data points based on the similarity to their closest neighbors in the training data.

pros
K Nearest Neighbors (KNN) ke fayde yeh hain ki yeh aasan hain इस्तعمال mein aur yeh bina kisi training ke complex data ko samjh sakte hain (K Nearest Neighbors (KNN) advantages are that they are easy to use and can understand complex data without any training).
I can definitely help you with that. In one sentence, K Nearest Neighbors (KNN) is a machine learning algorithm known for its simplicity, ease of use, and ability to handle various data types without requiring complex assumptions.

cons
K Nearest Neighbors (KNN) andar k data points ढूंढने में धीमा हो सकता है और बहुत सारे डेटा के लिए काम नहीं करता है। (K Nearest Neighbors (KNN) can be slow to find k data points and doesn't work well for very large datasets.)

working
KNN predicts by finding the most similar data points (its k neighbors) from the training data and mimicking their label (classification) or value (regression).

overfitting
Machine learning mein overfitting woh होता है जब आपका model training data को इतना अच्छे से सीख लेता है कि वो नया data handle नहीं कर पाता.
overfitting ek aisi samasya hai jismein model training data ko itna acchi tarah se fit karta hai ki woh unseen data par achcha perform nahi karta hai.**
Iska matlab hai ki model training data mein har pattern aur noise ko bhi sikh leta hai, jo unseen data mein nahi hote hain. Isliye, overfitted model unseen data par galat predictions karta hai.

Underfitting
machine learning mein aisa होता है जब model data को सीख नहीं पाता और नतीजा गलत predictions देता
Machine learning mein underfitting tab hota hai jab model bahut simple hota hai aur data ki complexities ko capture nahi kar pata hai, jisse training aur testing data dono par performance kharab ho jaati hai.

resampling technique
ka matlab hai kisi dataset se randomly sample nikalkar unka use karke naye data points generate karna.
Machine learning mein, resampling techniques existing data ke subsets create karti hain jisse model training aur evaluation mein aapko madad milti hai.

holding a validate
मशीन लर्निंग में, वलिडेशन डेटासेट वो जानकारी होती है जिसका इस्तेमाल ये चेक करने के लिए किया जाता है कि आपका मॉडल कितना अच्छा सीख रहा है

binary classification

Machine learning mein binary classification yeh batane ka tarika hai ki koi cheez do groups mein se kis group mein aati hai.
Machine learning mein binary classification yeh batane ka tarika hai ki koi cheez do groups mein se kis group mein aati hai.

generalized linear model (GLM)
ek aisa model hota hai jo dependent variable aur independent variables ke beech sambandh ko ek linear equation ke roop mein darshata hai, jismein dependent variable ka distribution ek specific family se hota hai, jaise ki normal, binomial, ya Poisson.
Machine learning me generalize linear model (GLM) ek aisa statistical model hai jo linear regression ko expand karta hai aur output distribution function ki variety badhata hai, jisse yeh non-normal data ke liye bhi kaam karta hai.

Logistic regression
is a machine learning method used to predict the probability of something happening, like whether an email is spam or not.
Machine learning mein, Logistic Regression ek aisa algorithm hai jo binary classification problems mein probability values ka use karke data ko classify karta hai.

sk learn
datasets provide karta h