Great Learning
POWER AHEAD

DATA DIGEST BITS & BYTES

FEBRUARY 2022 EDITION

What's Inside?

GREAT LEARNING SUCCESS STORY

Mahesh Venkatasubramanian
PGP-DSBA Alumnus

I am a 51-year-old sales professional who evolved from a door-to-door salesman to the Sales head, with a B.Sc. mathematics as my major in the early 90's. I would like to touch upon my personal background, before getting into what transformed me.

I had undergone a brain surgery in September 2019 when I was almost 50 years old, which was performed on my request and I knew my professional career had come to an end. I was unwilling to give up; hence, I planned to hone my skills further rather than wallowing in self-pity because of low self-esteem.

My options were to go ahead either with an Executive MBA from a premium institute or take up a few odd jobs as a sales consultant. It was divine providence that got me reconnected with my childhood friend, a senior professional in the HR department of a premier management institute in India (one of the top 5). He was categorical that under my current health condition and age, pursuing an executive management program will not add great value to my career or needs. He recommended I pursue a Data Science program with Great Lakes. My son, a software engineering student in a premium institute, also encouraged me to go ahead with the program, knowing that I am averse to coding. It was a great risk, as I had no savings, and most of my earnings were spent on health and other superfluous things.

I took the bold decision of joining the PGP-DSBA course from Great Learning. I passed out as a batch topper, excelling in all modules of the program. I had to skip one batch, which was advised by my program manager, who saw me operating more with an alarm clock, as I was still recuperating from post-surgery discomforts. By then, I resigned from a corporate company and joined another small organization with a lower remuneration, but more flexibility in work timings.

Today, many people asked me about my career as a data scientist. My answer is - I know I will get a good job, but I am not keen on pursuing that as my whole career. Instead, I took up the mission to train underprivileged kids as data scientists. I have received offers of visiting faculty from tier 2 management institutes, which I would love to experience, as I also have 3 decades of domain experience in sales. It is imperative for an MBA graduate to have an understanding of data science. This program played a critical role in my life as it catapulted my self-esteem and my self-confidence, enabling me to take care of my family and needs.

I have a lot more to say, however I would like to express my heartfelt gratitude to Great Learning, which not only formulated and put together a great curriculum, but also hired some of the most passionate faculty. My special thanks to some key people who played a vital role in my life - Professor Raghav Shaam (Faculty for Marketing and Retail Analytics), Professor Gurumoorthy Pattabhiraman (Faculty for ML and Statistics and my program manager Jenlyn Jude Miranda. I will be continuing my blogs on Great Learning at every stage, as I move closer to motivate other people.

Though the list of people to thank is a long one, I will always mention them in my future blogs as the association with Great Learning will continue for a long time. These blogs should be a motivation for all those who don't have any knowledge about coding, but are looking for a career transition in an in-demand domain.

DISCOVER

Machine Learning - The Next Big Thing?

Machine Learning is an application of Artificial Intelligence (AI) that allows computer programs to progressively learn and improvise from their experience with the data. Therefore, Machine Learning matters; as it shapes your ease of living or decision-making. It has been integrated into our daily life so deeply, that you will most likely not notice its application. For example - the active filtering of your spam messages by Gmail.

Some of the most prolific Machine Learning users are in the banking and financial industry. For example - HDFC Bank has begun rolling out its technology stack with ML and AI. Major e-commerce platform Flipkart implements over 60 Machine Learning models to generate insights – "how a sale is going, which deals are working or not working, at which point are customers dropping off" and so on. In the HR sector - Aspiring Minds has an assessment-based job search platform for adding value to merit-based recruitment.

LEARN MORE

Tableau Tutorial: A Beginners Guide to Tableau Desktop

Data visualization is a graphical representation that explains the relationship and behavioral patterns in data. Data visualization is an integral part of analytics, including Data Science. Data visualization compresses a huge amount of virtual data, both big and small, that is difficult for the human brain to grasp and represent. To determine the appropriate analytical tool for business it is necessary to consider the following aspects:

- Amount of data being processed
- Frequency and type of reports and structure
- Cost of the solution to determine the right analytical tools

To know more in detail about hands-on Tableau, let's dive deep into it.

LEARN MORE

Real-time Face Detection using OpenCV

In this article, we are going to find out how to detect faces in real-time using OpenCV. We are also going to see how to make a custom mask detector using Tensorflow and Keras.

The goal of face detection is to determine if there are any faces in the image or video. If multiple faces are present, each face is enclosed by a bounding box and therefore, we know the location of the faces. There are two main approaches for Face Detection:

- Feature Based Approach
- Image Based Approach

LEARN MORE

SOCIAL MEDIA BULLETIN

With Great Learning being mentioned in coveted newspapers all the time - like Times of India, Economic Times, Financial Express, Business today and other websites for its high-quality learning and the upskilling opportunities provided to thousands of learners - we have become quite popular both domestically and internationally for our wide range of quality programs. With thousands of learners finishing various programs with us every year, we have been getting a lot of attention on various social media platforms like LinkedIn, Instagram, and Facebook. These highly satisfied learners have been vocal in sharing their learning journey with us.

Ramya Jagannathan • 2nd
Analytics Modeler Machine Learning,Ford Motor Company
1mo • Edited •

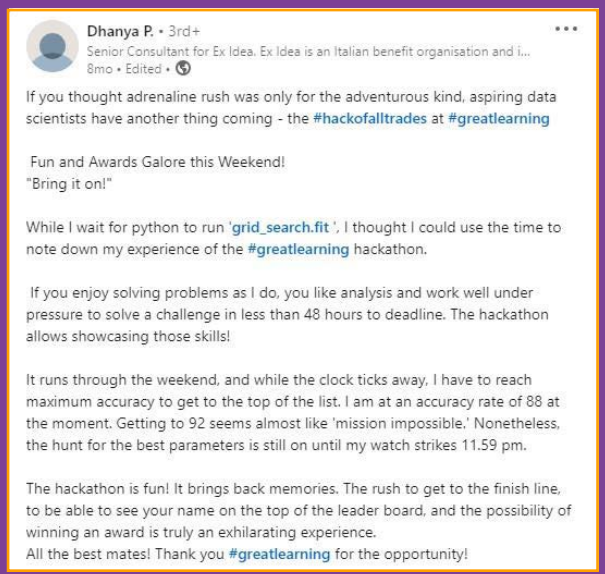
Hello connections!

Here is my first post in 2022 and I am immensely proud of it. Happy to share that I have completed my PGPDSBA course at Great Learning
A wonderful learning journey and I really mean it. Thanks to my mentors for helping me understanding different concepts and a big applause to program manager Sonali Jangwal Maan for being extremely helpful and supportive and for always being there. Thanks to my friends who travelled with me and made this journey more fun.

I am also extremely happy to share the link to view my digital certificate provided by University of Texas .

https://vrfy.digital/LYCMW

#certificateofcompletion #mygreatlearning #feelingproud #powerahead



Dhanya P. • 3rd+
Senior Consultant for Ex Idea. Ex Idea is an Italian benefit organisation and i...
8mo • Edited •

If you thought adrenaline rush was only for the adventurous kind, aspiring data scientists have another thing coming - the #hackofalltrades at #greatlearning

Fun and Awards Galore this Weekend!
"Bring it on!"

While I wait for python to run 'grid_search.fit ', I thought I could use the time to note down my experience of the #greatlearning hackathon.

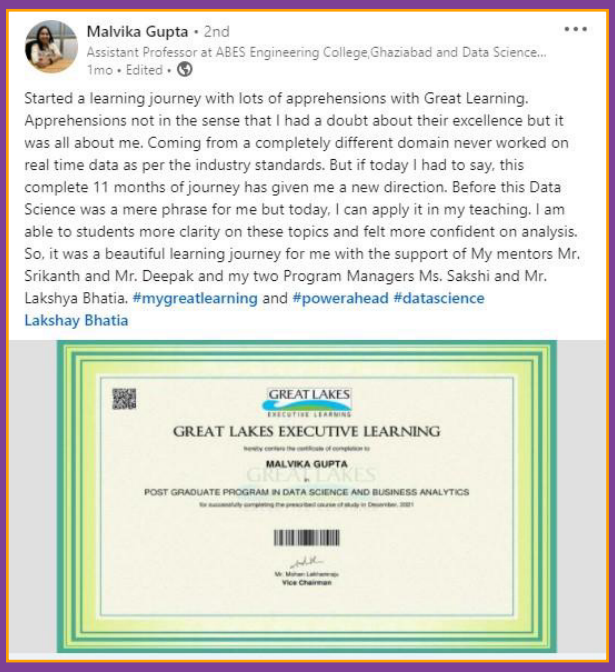
If you enjoy solving problems as I do, you like analysis and work well under pressure to solve a challenge in less than 48 hours to deadline. The hackathon allows showcasing those skills!

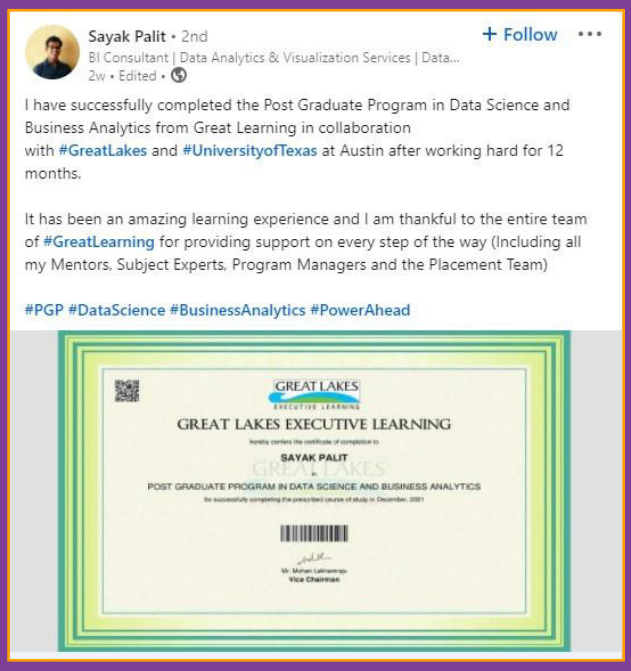
It runs through the weekend, and while the clock ticks away, I have to reach maximum accuracy to get to the top of the list. I am at an accuracy rate of 88 at the moment. Getting to 92 seems almost like 'mission impossible.' Nonetheless, the hunt for the best parameters is still on until my watch strikes 11.59 pm.

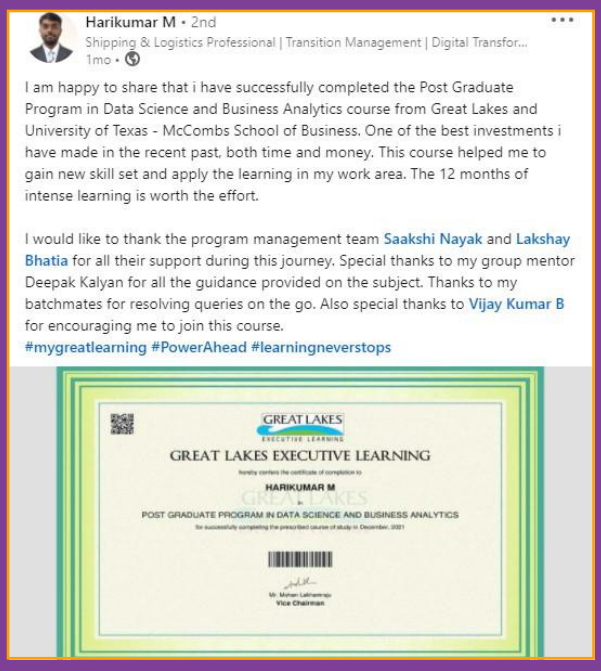
The hackathon is fun! It brings back memories. The rush to get to the finish line, to be able to see your name on the top of the leader board, and the possibility of winning an award is truly an exhilarating experience.
All the best mates! Thank you #greatlearning for the opportunity!

The key highlights of these posts are the superior quality content and interactive mentoring sessions. The learners have been highly appreciative of the various activities which are planned for them, thereby making their upskilling journey more fun and interactive.

The learners have been practicing “Gurudakshina” - an activity that involves thanking their mentors in the learning journey. These activities not only make the journey fun, but also help learners bond with each other. All in all, our high-end quality programs have the wings of strength through these incredible social media posts.







SPOTLIGHT

ANALYZING CREDIT RISK OF OIL AND GAS COMPANIES

This edition of spotlight features the most important factors affecting the credit risk of oil and gas companies. Read and learn how Shamba Datta, Daisy Mary, Vinaya Pagadala, Anbu Jacob, Prabu Arumugam of the PGP-DSBA, Chennai - March 2019 batch helped investors identify the risk potential of their loan in oil and gas companies. They utilized the financial data of the past 15 years and the study was later published in an AICTE sponsored ICDML-2020 conference.

The initial step of this study was collecting the financial data of 8 international and publicly listed oil and gas companies over Quarter 1, 2005 till Quarter 3, 2019. The quarterly data was analyzed, after which the companies were classified into 3 different risk categories - low risk, medium risk and high risk - by using Moody's Credit Rating technique. After studying the numbers, the researchers noticed that there was a strong relationship between a company's credit risk and other financial parameters. The project was aimed at classifying the credit risk using a simple credit scoring methodology utilizing minimum resources. The results obtained indicated that the data was skewed for most of the parameters. In conclusion, the study stated that the risk is directly related to the profitability ratio of the organization. The strategy was to compare the financial performance of the company for different periods and its partners. This study helped in determining the improvement or deterioration in a company's credit risk.

This approach helped identify the early trouble signs in a company's credit risk that could affect the shareholders, the management team and the lending agencies, thereby prompting organizations to take decisive actions.

CAREER SECTION

Let us emphasize on the need for a good resume and how to make one.

A resume is a reflection of your achievements and profession, and should have an everlasting impression on the employer/talent acquisition team.

According to the employers/acquisition teams:
- 80% of the resumes are misleading
- 70% of resumes have either wrong email or phone numbers

Nowadays, most resumes are scanned through the ATS (Application Tracking System) and therefore, attention-grabbing keywords are a must.

7 Steps to build DS and ML Resumes

1. Layout & Format

DO's
- The lesser the better, the length should be a maximum of two pages (one page is ideal).
- Highlight keywords, like in-demand technologies and industry skill sets.
- A PDF format is desirable. However, word files can be used as an alternative.
- A maximum of two fonts can be used, but uniformity must be maintained.

DON'Ts
- Refrain from being extremely artistic; standard templates work well.
- A photograph is not required.
- Information such as nationality and marital status is not required.
- Unnecessary visualization should be avoided.

2. Cover Letter

- Good to have this when applying for jobs in research or teaching fields.

  In the Data Science industry, this is not essential.

  If at all required, it should:

1. Contain achievements related to the role.
2. Highlight skills and experience needed by an employer.
3. Show genuine enthusiasm towards the role.
4. Significant achievements from the previous role.
5. Tell the recruiter why you are the best suitable candidate for the job.

3. Introduction Sections

DO's
- Should always be on top.
- Full name - as on passport or ID proof.
- Valid contacts (mail ID + phone number).
- Kaggle/GitHub/LinkedIn profile link.

DON'Ts
- Complete address.
- Languages – only if required for the role.
- Logos/visuals.
- Write-up/elevator pitch–optional.

4.Professional Summar

DO's

- A 3-4 line write-up about your professional experience – strengths and skills.
- Oriented more towards technical (DS and ML) than general content.
- Using keywords according to the job description.

DON'Ts

- Career objective/seeking opportunity – not a distinctive factor.
- Stick to the points which align to the applied job role.

5. Technical Skills & Expertise

DO's

- Based totally on the job description – use essential keywords.
- Key skills and expertise required for the role.
- Structured view.
- Only one technical and business skills section.

DON'Ts

- Too many or too less techniques.
- Self-rating
- MS Office
- Too many soft skills for a tech role.

6. Work Experiences

DO's

- Experienced professionals – Relevant projects on the top.
- Freshers – capstone project, academic projects, hackathons or practice problems – prioritize based on the JD.
- Structured View –
  a) Problem statement
  b) Approach/solution
  c) Outcome/business impact
  d) Implementation/deployment
  e) Include links (if available)

DON'Ts

- General skills like MS Office.
- Random visuals.
- Long project descriptions.

7. Academic or Personal Projects

DO's

- Experienced professionals who are not from the Data Science and Machine Learning domain – capstone project, academic projects, hackathons or practice problems – relevant based on JD.
- Maximum 3-4 projects.
- For a very senior profile – do not focus on individual projects. Instead, focus on the initiatives related to Data Science and Machine Learning
- Includes links (if available).

DON'Ts

- Many common projects like - basic analytics, non-technical projects etc.
- Long descriptions.
- Repetitive content/projects.

WHAT'S NEW?

DARWIN integration on LinkedIn would suffice the needs of Data Scientists

LinkedIn, as we all know, is the largest professional network on the planet, generating massive amounts of high-quality data. For Exploratory Data Analysis (EDA) and Visualization, data scientists and AI developers have been employing a variety of tools to interact with data via several query and storage engines. As a result, a single "one-stop" Data Science platform was required to integrate and fulfill many requests.

DARWIN's focus on use-cases is very similar to those serviced by other industry-leading data science platforms. It employs the Jupyter ecosystem, but it goes beyond Jupyter notebooks to address the demands of LinkedIn's data scientists and AI engineers. And this is where DARWIN comes into action.

DARWIN was built to address the needs of all LinkedIn data creators and users, not just data scientists and AI engineers with varied skill sets. It was vital to establish the different personas that DARWIN needed to serve and address. Jupyter notebooks are often used for data visualization and analysis. AI specialists employ a variety of machine learning libraries to train and evaluate different ML algorithms, including GDMix, XGBoost, and TensorFlow.

Big Data is changing fashion trends across the globe

Ralph Lauren uses big data in his business! Their polo shirts have sensors connected to them that track the consumers' fitness levels. Buyers may track their health with the use of a smartphone app every time they wear a Ralph Lauren polo. David Lauren says, "I want to be able to gather this [biometric] information in a boardroom or from a baby in a crib. We'll find new needs, and we're just at the beginning". WGSN, a fashion firm established in the United States, uses big data to forecast future fashion trends. It also provides information about your customers.

The fashion business collects a vast amount of consumer data from a variety of sources in order to develop their products, including fashion taste, needs, interests and preferences. Editd, a big data platform for retailers, helps fashion outlets determine clothing and accessory prices, and sell their products more effectively. With data streams from several social media networks, the programme gauges customer sentiment.

As a result, big data is already making an impact. In the fashion and design industry, it is being used to understand customer preferences and assist fashion houses in marketing their products.

NFT Terminal has launched a platform for real-time data analytics

After the successful launch of their initial service, which indicates that more than 100 communities can do rare searches, NFT Terminal has opened its beta release with over 730,000 users in its first month. NFT Terminal is now available on mobile and desktop applications and the beta version is available for free.

NFT Terminal includes the following features:

- Real-time data insights from well-known whales and influencers.
- Buys, sales, stakes, and transfers are all available in real-time.
- In 1-hour, 4-hour, and 12-hour intervals, rankings for the most traded NFTs, recently acquired NFTs, sold-out NFTs and the most minted NFTs among whales and influencers are available.

DATA SCIENCE AT WORK

Humera Yasmeen, Trichy
June 2021

Find out how Humera utilized Data Science to improve the safety aspects in the aviation industry.

Humera Yasmeen works as a Senior Technical Lead in software development under an Agile project in the aviation sector for flight safety management. She has an experience of over 10 years and has worked across various domains such as banking and insurance, healthcare, hospitality, automation, and aviation. She is proficient in front-end and back-end development. In her current organization, she got a chance to work on a project on safety management for flight and risk assessment.

It is a general practice to collect information (such as health conditions, sleep duration, flight conditions, weather conditions, flight operational status etc.) from crew members to ensure safety in the aviation industry. In case of an accident, a detailed report is submitted to the concerned departments, along with their inputs. The data that was shared by the team had to be collected and mapped in the form of a risk matrix to predict potential risks. The comments had to be read manually for deciding the action to be taken thereby, making it a tedious job which further delayed analysis.

USA flight data for the last 10 years was collected, and that included the number of aircraft accidents that took place within that timespan as well as the forms that were filled out before the flight would take off.

Time series forecasting was used to clean the codes, remove unwanted columns and analyze the trends in the flight accidents. In addition, Natural Language Processing (NLP) was used to categorize the accidents as environmental issues, technical issues, the action of the pilot, etc. The analysis helped Humera understand how flight conditions have implications on the crash. The data also analyzed the psychological patterns of the pilot and evaluated how they impacted his/her decision-making capabilities. This approach involved various techniques (Logistic Regression, Random Forest, Multinomial Naive Bayes, Support Vector Classifier, Decision Tree Classifier, K Nearest Neighbor, Gaussian Naïve, Bayes Theorem) and programming languages (such as Python and NumPy), and flask for model deployment in local machines.

Humera was able to generate a model using the data collected before a flight take-off and was able to predict the potential risk associated with the journey. The information obtained from the personnel was processed using NLP and helped in identifying issues that were redirected to the concerned aviation teams automatically. There has been a considerable decrease in the time taken for take-off after automating the process. Earlier, flight accidents and the First Information Report (FIR) required extensive human intervention to close the issue. It is believed that this prototype will reduce human intervention in risk assessment as well as the total number of hours taken to ensure the desired take-off.

Humera learned about the difficulties in data collection, filtering the initial data, grouping it and arriving at a conclusion. Overall, 10 models were generated which were compared with their residuals. With NLP, Humera was able to prepare the data, tokenize and segregate the data as per the inputs received. She also learned how to deploy models with Flask and predict the risks associated.

Mrinal Roy, Mumbai

May 2021

Read this story to find out how Mrinal and his team formulated an approach to drive the sale of safety equipment during the pandemic.

Mrinal Roy works as a Product Specialist at a manufacturing firm that produces Personnel Protection Equipment (PPE) including helmets, eyewear and shoes. He is responsible for driving sales for various safety equipment like eyewear, face masks and capital equipment such as flammable cabinets.

During the pandemic, his team was facing many challenges. In his organization, like most businesses, the sales team used to interact with the customers directly and understand their requirements, which helped the Sales team to pitch the right products. When the pandemic started in March 2020, pitching the products over calls and online interactions became tough as the customers did not show much interest.

To address this issue, Mrinal and his team pulled out the data that consisted of purchase patterns over the last 3 years. The data helped them identify the end-users, distributors and understand their preferences. It consisted of 2500 uniquely billed customers. The technique of Exploratory Data Analysis (EDA) was used featuring univariate, bivariate and multivariate analysis, which helped his team learn more about the purchase patterns.

By conducting EDA, Mrinal and his team were able to generate details for the purchase patterns and identified the products in demand. The details included the month of purchase, the types of products supplied and the customer preference. They also built product packages for recommendation through the online mode. These details helped Mrinal in organizing his calendar schedule for sending out emails, calling potential customers, and arranging webinar meetings with them. The approach helped in ensuring that the right products are delivered to the right customer, along with proper scheduling of products over emails, webinars and telephonic conversations. The data helped the team in analysis and detailed reports. The strategy boosted the number of inquiries generated, which saw a considerable increase of 10 % for the first month and 35% increase at the end of the 5th month. Overall, his team was able to regain 42% of the customers by employing this approach.

THAT'S A GOOD QUESTION!

In this edition,
we will learn the

difference between Collaborative filtering and Content-based filtering.

Mr. Rakesh Lakkala says:

What is a recommendation system?
A recommendation system is the ability of a machine to predict the user's preference based on the user's profile and his historical preferences data. Recommendation systems are widely used in any B2C market to drive their sales and to improve customer interaction. Some customers who use recommendation systems are Amazon, Flipkart, Netflix, Hotstar, YouTube etc.

2 types of recommendation systems:

1. Collaborative Filtering
2. Content-based Filtering

Collaborative Filtering:
This method uses past interactions between users and items to produce new recommendations. Here, we try to find what similar users will like and those users with similar interests are clustered. Recommendations are provided to the user based on the cluster he/she belongs to. This is a memory-based approach. For example - if users A, B and C belong to one cluster having similar interests, then another user D having similar interests will also fall in the same cluster, Therefore, the recommendations to D are given based on the preferences of A, B and C. This collaborative filtering technique technique is called Nearest Neighbourhood algorithm. Collaborative filtering can be user-based or item-based. The main challenge in this technique is that recommendations for unseen items in training cannot be conducted and is called a cold start problem.

Content-based Filtering:
This method uses customer features and items, to recommend products to users. In this method, a utility matrix is used to find the relationship between users and products based on the features. A value called degree of preference is assigned for each item user pair. However, this does not depend on historical data. When a user adds items to his/her cart or selects a few items for his watch list, automatic recommendations are given based on the selected items, thus avoiding the cold start problem. Recommendations using this method are more relevant to users compared to collaborative filtering. The main challenge in this method is to add feature tags every time an item is added, as adding incorrect tags results in incorrect recommendations. Hence, for better recommendations, better subject matter experts are needed.

Mr. Saurav Sidhwani says:

Both Collaborative and Content-based filtering comes under the recommender system. Recommender systems are nothing but algorithms that provide consumers with relevant recommendations. “People you might know” on Facebook, Instagram and LinkedIn, “People also watched this” on Netflix, Prime, Zee5, “People who bought this also bought this” on Flipkart, Myntra etc. are nothing but the applications of recommender systems.

Collaborative Filtering:
This type of filtering relies on the data provided by different users, either by giving a rating or clicking on the link of an item. In this case, we save the details of users - like the different items they want to choose or like and based on the similarities between the likes of different users, we make a prediction of what the users may like based on what similar users did. Everything will be summed up and items with the highest score will be recommended.

For example: There are 4 users and all of them like to see thriller movies and table A, B, C, D, E, F, H, I and J correspond to movie names.

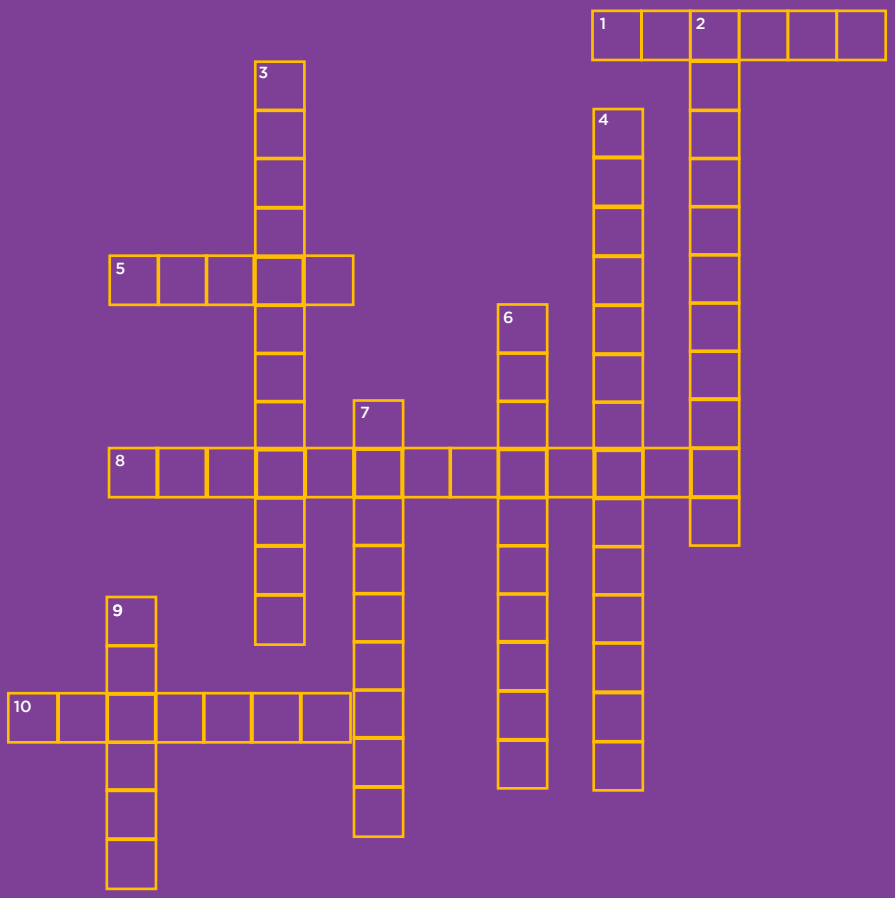
User 1	User 2	User 3	User 4
A	A	A	
B	B	B	
	C	C	C
	D	F	I
	E	H	J

User 1 has watched only movies A and B and has not watched any other movies. Using collaborative filtering, this data will be matched with other users (users 2,3,4) and movies liked by the other users will be recommended (probably C in this case).

Content-based Filtering:
In this type of filtering, the system uses the features and likes of the users in order to recommend the things that the user might like.

For example: I am a big fan of Marvel movies and often watch sci-fi movies on the internet, and you all might know that google saves all our search history. Therefore, when I install a new app and that app reads the data from my search history, it automatically starts recommending movies from the sci-fi genre that would be ideal for me to watch.

DATA SCIENCE CROSSWORD



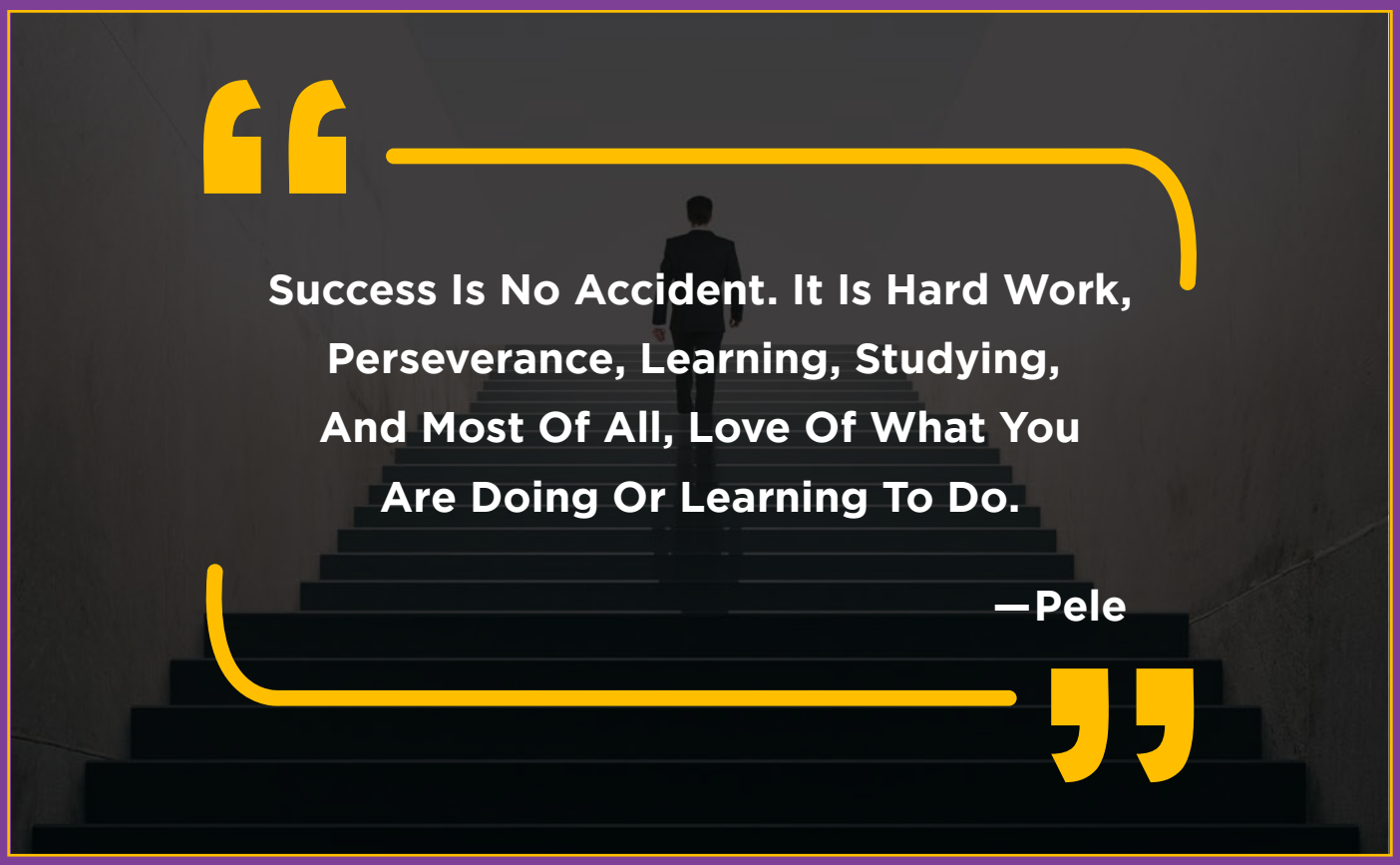
ACROSS

1. Open-source software framework that allows data scientists to process big data using clusters of hardware running simple programs.
5. A methodology cribbed from software development that now sees application in many areas of business.
8. The measure of the statistical accuracy of an estimate decreases with the larger data set.
10. Data about data, or data attached to other data.

DOWN

2. Values, behavior, and norms shared by most individuals within an organization regarding data-related issues.
3. Branch of machine learning that attempts to mirror the neurons and neural networks associated with thinking in human beings.
4. A time series model that uses observations from previous time steps as input to a regression equation to predict the value.
6. A category that can be used to arrange data by facts and measures for data dicing (grouping) and slicing (filtering) purposes.
7. Provides at-a-glance statistical analysis and historical trends of an organization's KPI's, using graphical representations.
9. Code-sharing and publishing service, as well as a community for developers, offering repositories and new accounts.

LEARNING BIRD CHIRPS



THE EDITORIAL TEAM

Garima Chamoli

Srijan Purang

Abhirup Dey

Mack Donald