Great Learning
POWER AHEAD

BITS & BYTES

JULY
2022 EDITION

WHAT'S INSIDE?

LEADERSHIP SPEAKS

ABHISHEK GUPTA

Associate Vice President - Products, Great Learning

1. What is one characteristic that you believe every leader should possess?

One characteristic that every leader should possess is the willingness to take care of the team. There should be trust between the leader and his/her team and a leader should be able to make human connections with the team.

2. Talk about a leader that inspires you and why?

Inspiration can come from anywhere. On the professional front, I have been fortunate to interact and work with many able leaders that had a positive rub-off on me. In addition, there have been many successful personalities whose phenomenal work and contribution towards the betterment of society influenced me. Recently, I have been following Yoga guru Baba Ramdev, who has been instrumental in spreading the awareness of Yoga and Pranayam for a healthy lifestyle and has been taking several initiatives towards keeping alive the rich heritage of best practices from our ancestral days. Covid and its after-effects have made us all more conscious about our health and lifestyle. As a leader, he has been able to influence millions of people across the world to practice Yoga and meditation as well as adopt Ayurveda and naturopathy to live a healthy life and inspired many people in the world by living a disciplined life.

3. What advice would you give someone going into a leadership position for the first time?

Leadership is about the willingness to take care of others and have empathy. A leader thinks about others' interests before making decisions. One should start thinking about how the team members could succeed. The time to take individual credit is over and now it is time to make others shine. The responsibility of a leader is to ensure every team member can do their job well. There are a few other things which an individual should start practising before transitioning from an individual contributor to a leadership position:

- The specific skills or expertise that you had as an individual contributor may not be sufficient now, what got you here will not get you there. One has to start looking at the bigger picture and see how the expertise of the entire team could be leveraged to make things happen.
- You don't need to know everything and have answers to all the problems. Your job is to ask thoughtful questions to help the team find the answers. A leader should empower the team by sharing all the relevant information. Also, how you ask a question is more important than what you ask.
- Become a good listener, listen to understand and not respond.

4. How do you help a new employee understand the culture of your organization?

The culture of an organization is defined by the beliefs, attitudes and behaviour of its employees in their day-to-day life. To make a new employee understand the organization's culture:

- It is important to tell them the story about how the company has evolved over the years and how the employees across all levels have been behaving and repetitively doing things in a certain manner, which is aligned with the organization's values.
- Lead by example - Demonstrate the values and company culture during the entire onboarding experience and every subsequent interaction with the new employee i.e. during the interview process or interactions before the joining date.
- Encourage the new employee to ask questions and assign a buddy (preferably those who have been with the organization for a long time and have imbibed the culture) to reach out for any clarification.

5. What's the best book/movie/series you've read/ watched this year? Do share takeaways.

I recently read an amazing book titled 'Influence' by Dr. Robert D. Cialdini. It is a masterpiece by Dr. Cialdini, where he has shared his learning from extensive 35 years of research on the minds of the people and their general psychology.

This book aims to demystify the psychology behind how humans are persuaded. There are six universal principles, the knowledge of which can help us persuade others:

- When someone does something for you, you are obliged to return that favour.
- Once we have made a choice or taken a stand, we will encounter personal and interpersonal pressure to behave consistently with that commitment.
- When you are unsure how to behave, you see how others are behaving to guide your action. (We are more likely to do this when we view others as similar to ourselves.)
- We prefer to say yes to requests of someone we know and like.
- We are more likely to be persuaded about something when it is coming from a trusted source.
- When something is available in a limited quantity, we are more likely to develop the desire to possess it.

GREAT LEARNING JOURNEY

PRABAKARAN
PGP-AIML ALUMNUS

I work as a Technical Lead Developer at 3i Infotech. I have around 5+ years of professional IT experience in net development and utilised Microsoft tools such as Azure, Asp.net core, .net core, Azure DevOps, and Cloud Automation. Before joining the PGP-AIML Course, I was working as a Senior Software Developer at TCS.

I feel the biggest challenge is managing time. I found it difficult to manage time with the course and work parallelly. I wanted to do an MBA from Great Lakes initially, but then planned to go for the PGP-AIML program. I had heard a lot about this program and obtained positive feedback from the alumni.

The mentored learning sessions were essential to address the doubts and even helped me when I didn't go through the weekly content. These sessions emphasized the need for finishing the weekly content on time. Sometimes the program content seemed very theoretical and required further examples to get a hold of the concepts. The mentor connected the dots and ensured a firm grip on the concepts learnt. It helped us in relating the technology that has evolved over time.

The lectures were a mix of in-depth discussions and clear explanations of concepts. All the queries were addressed by the mentor effectively. We were introduced to a lot of real-life problems and hands-on projects. The mentor was supportive throughout and shared lots of reference materials to facilitate our learning. The hard part for me was working on the project but I utilised the mentor's teachings to decode the problem statement.

The skill sets that I learned have ensured that I am well equipped and capable of using multiple AI platforms. Currently, I'm using Azure AI. I hope this can further help me in making a transition from the current role to an individual contributor role in Data Science.

My advice for people seeking to start afresh in AI - spend enough time learning the new concepts. It's a patience game, once you start learning it, you will start enjoying it.

GREAT LEARNING JOURNEY

ANNA THOMAS

PGP-DSBA ALUMNUS

I was working as a manual tester in Tata Consultancy Services. I quit my organization in 2019 as I wanted to be with my children. I have worked with mainly three clients i.e. AMEX, Bank of Montreal and Citibank. My main work involved creating test cases, getting relevant data and validating various user scopes based on the scenarios explained by the business and have worked both in Waterfall and Agile methodologies. It was a wonderful experience working with TCS but I felt there was scope for improvement both in terms of career growth and transition.

I wanted to grow further in my career and was always interested in knowing how the testing data was utilised by the organization. I was always intrigued by the way the business analysts I had worked with, explained concepts to our stakeholders with the data available. I wanted to learn how they ensured seamless connections between the business and developers. My biggest professional challenge was staying updated with the latest technology and trends.
I had tried various centres and found Great Learning to be more reliable as they offered a wholesome package and did not have any apprehensions regarding the online mode as I could learn easily from the comfort of my home.

It was an awesome learning experience as I had wonderful mentors and professors who guided us through online mode.

My program manager was very helpful and helped me out in various emergencies and responded quickly whenever needed. The schedules were crafted in a way that we could understand and at the same time devote time to our classes properly. Our mentor was good at explaining all the concepts and clarified all the doubts meticulously. After the class, he would also send us the mentor session material which served as a refresher material for the completed subject. This helped us understand the program in general. The mentor also shared many pointers to follow while coding. The program material was available in abundance and that helped me understand all the subjects properly.

I have regained my confidence and also feel that I am now up to date with the latest technologies, thanks to this program. I am yet to join the workforce as I am planning to try for a new job, once I complete the electives we have been given in the program.

My advice to people who are passionate about understanding how data works and how it's useful - the PGP-AIML is an excellent choice. It is worth the journey and we have a lot to gain from it.

SPOTLIGHT

DETECTING COMPLICATIONS IN X-RAY AND CT SCAN REPORTS USING AI APPLICATIONS

This project was submitted by Krishna Kumar along with his peers from the AIML Online 2020 batch. The idea was realised when Krishna was left fascinated after listening to Sundar Pichai's address at a Google conference in 2018. He was impressed how AI can help in eye care, detect the patient's conditions and even predict other ailments that the patients were facing but were unaware of!

Their project focused on analysing the prospects of disease identification by examining X-ray reports using AI. As a current practice, the report is studied by a specialised doctor (Radiologist) before starting the treatment. The project does not talk about replacing the radiologist but rather augmenting his capabilities using AI. This could prove to be a major breakthrough in remote areas wherein there are limited resources available for treatment or expert radiologists are sparse. This system will significantly reduce the existing load and scale their expertise.

Earlier in Apr-May 2020, Krishna and the team decided to do a project on how to apply AI on X-Ray or CT scan images. After a week and in discussion with their Professor, Dr Narayan, they came up with the idea of doing something worthwhile to tackle the COVID pandemic. Referring to the research papers shared by their professor, the team was excited to apply AI in this emerging area and was curious to note the findings.

Being full-time employees, they had time constraints. In addition to it, they faced computational challenges using Google Colab and AI being a heavy application had its own restrictions. The choice of algorithm and proper direction was facilitated with the help of the mentor who motivated them to experiment. The deficiency of data in the COVID X-rays experiment was another roadblock in their endeavours for which they studied previous papers on COVID research. At least three iterations were conducted after going through the published data using which they derived better results. Once the team started with a very lean data set for COVID, they devised algorithms to train with imbalanced data sets. These strategies help to fix these major roadblocks.

The team was able to highlight the unexplored algorithms useful in detecting lung pathologies accurately with fewer computing resources (DarkNet-53). The algorithm can accurately detect pathology even with very few samples for the class of interest (Ex- COVID). Also, this algorithm can be trained on either X-Ray or CT-scan images with potentially. This paper also gives a comparative analysis of various transfer learning algorithms and how the DarkNet-53 outperforms other algorithms with the same dataset.

This research paper was accepted at the CD-MAKE International conference in Austria-Graz and the team received lots of accolades for their efforts. The attendees were curious to know if this is being deployed by any startup or company. This project has emphasized the extended capabilities of AI.

DISCOVER

NO CODE AI IN 2022: MAKING ARTIFICIAL INTELLIGENCE ACCESSIBLE

As technology continues to impact human lives, the role of artificial intelligence (AI) is getting increasingly important. From improvement in healthcare services to engaging customer experience for an enterprise, AI has made our lives easier. However, developing AI tools for a specific industry requires a dedicated team of specialised developers who work along with domain experts to make it possible.
This is where no code AI tools come into the picture. It has enabled domain experts, businesses, and organizations to develop applications without any prior coding knowledge. Simply put, no-code AI makes AI more accessible and affordable to be used by everyone.

To read more about it, access the given link:
https://www.mygreatlearning.com/blog/making-artificial-intelligence-accessible/

NO CODE AI: APPLICATIONS ACROSS INDUSTRIES

No-code AI technologies are considered to be the next big thing in the digital world. It has brought sustainability and affordability to the expensive AI tools that, to date, were accessible by limited industries. The no-code AI approach eliminates the need for coding knowledge to implement artificial intelligence in businesses. These platforms enable non-engineers to apply AI and ML to classify and analyze data thereby streamlining the decision-making process.
The no-code AI expands its applicability, bringing AI to small enterprises that were earlier finding it expensive to use.

To know more about the industry-wise applications, please visit:

https://www.mygreatlearning.com/blog/applications-across-industries/

NO CODE AI: HOW IT DIFFERS FROM LOW-CODE AI AND AUTO ML

Demand for integrative AI has increased across sectors, leading to the rise in utilisation of tools such as AutoML, low-code AI, and no-code AI. No-code AI is a relatively newer concept that is aimed at democratizing the use of artificial intelligence. The technology has enabled users with no technical knowledge and experience to design applications and websites without writing any codes. The no-code AI platforms offer a drag-and-drop interface to deploy AI and machine learning (ML) models. It helps in testing ideas and building new projects and products quickly.
Similar to no-code AI are Auto ML and low-code AI. These platforms enable individuals to develop AI solutions without much investment and formal training in software development. They differ with regard to critical components and features.

To learn more, please visit:

https://www.mygreatlearning.com/blog/how-data-science-is-reshaping-the-future/

THAT'S A GOOD QUESTION!

In this edition, we will be focusing on what a decision tree is and why it is prone to overfitting.

Sridhar Anchoori says - A decision tree is a tree-like structure with multiple branches evolved based on the patterns in a given data set. This is one of the simplest yet highly effective supervised classification algorithms used for decision-making. The decision tree is used in any industry sector as the approach used in the decision is very simple and can be extended to any problem or scenario.

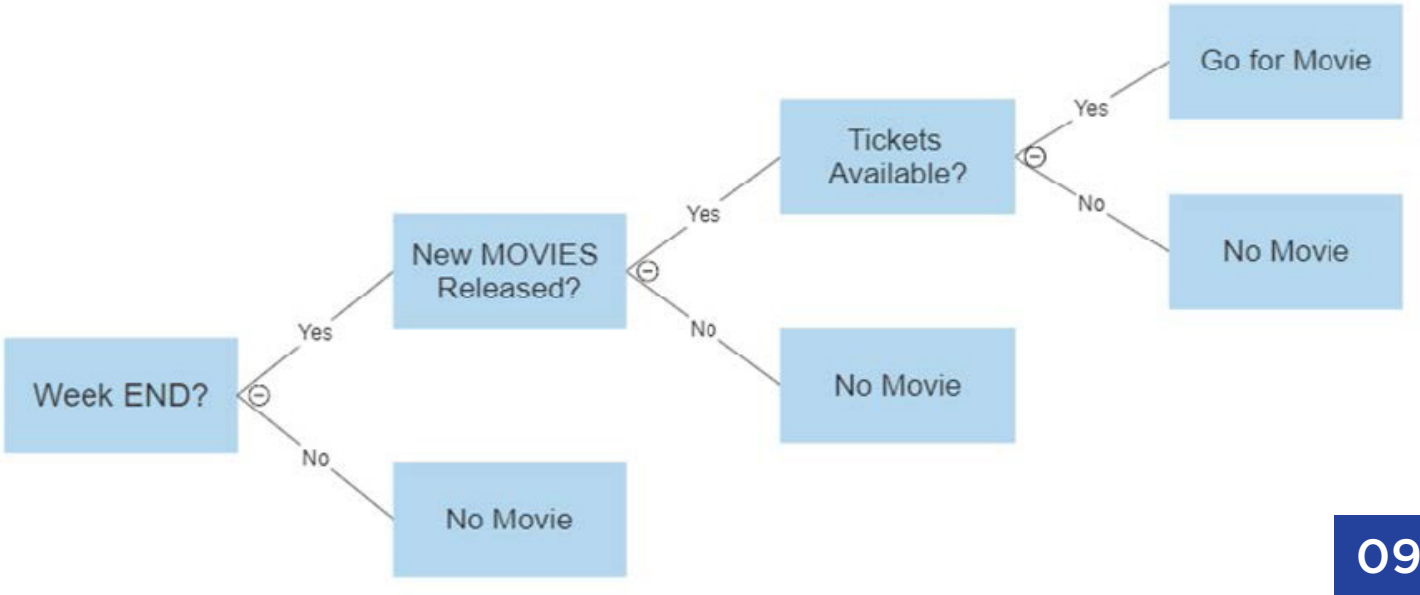
A decision tree contains multiple nodes and branches evolved using different questions and options. Each node in a decision tree is an independent variable with a condition and each branch is the resulting outcome. A decision tree can be applied for both categorical and classification predictions and is more effective in handling the nonlinear patterns in the data sets.

Let us take a simple example to understand a decision tree. For example, I want to decide whether to go for a movie tomorrow.

Before we understand why the decision tree is prone to overfitting, let us try to define it. Overfitting is where the model performance is significantly higher for training the data set whereas the model performance is significantly lower in the test data set. The model does not generalize well in case of overfitting and tries to perfectly fit all the samples in the training data set.

It is learnt that a decision tree uses nodes and branches to develop the model. While generating the tree there is a higher possibility that the model tries to create a greater number of nodes and branches to fit all the samples in the training data set. This increases the depth of the tree leading to overfitting. The decision tree uses purity indicators (i.e., Gini or Entropy) to find out the best variable to split, the main objective of these indicators is to find a node that is pure to be used for the splitting. The decision tree does a greedy search while selecting the pure nodes and this is also one of the reasons for overfitting.

Pruning technique is used to avoid overfitting and limit the growth of the decision tree. The main objective of pruning is to remove the number of branches that split on features with low importance. Pruning can be achieved by controlling the depth of the tree using the number of samples before or after the split, low impurity gains and maximum leaf nodes.



Jayveer Nanda says - Decision tree is a supervised learning technique that can be used both for classification and regression problems. It is a non-parametric approach that predicts the value of a target variable by learning simple decision rules that are inferred from the data features.

The decision tree uses a feature to split up at every node and that feature is selected based on various metrics like information gain, entropy or the gini value. At every step, that the variable is selected that has the maximum information gain or the minimum entropy or impurity.

The idea behind the decision tree is to use this feature to split up the tree in such a way that on the left-hand side the values are smaller and on the right-hand side the values are higher. For e.g., let's say we have an income variable and we have used that to split up the trees in such a way that on the left side of the tree the income is less than 50k and on the right, we have greater than 50k.

A decision tree exists in the 'sklearn.tree library' in python and by default doesn't specify any max depth parameter or any other parameter like min samples leaf etc. When you create a decision tree model, python keeps on splitting the data so that every leaf node is pure.

Since there are no parameters specified by default in the decision tree, especially the max depth parameter, it keeps on learning from the training data and creates a tree which has an infinite max depth and due to this, it often results in an overfit model i.e., a model which performs good on training data and bad on the test data.

To overcome this, we prune the decision trees by specifying the parameters which we find out by running 'gridsearchcv'.

INDUSTRY TRENDS

In this section of the newsletter, we will be discussing the various job opportunities available to data science aspirants and their corresponding roles and responsibilities.

Somak Sengupta, an industry expert, shares his perspectives about the skill requirements for people seeking jobs in the field of data science. He will also help us learn the basis for the naming conventions for various job roles.

The title 'Data Scientist' has been discussed widely mainly because of the ambiguity around what the role entails. The job descriptions are much clearer and the profiles are starting to normalize. This shift represents the maturity that the craft is achievable. Let us learn how the roles have been defined as per industry relevance.

DATA ANALYST / PRODUCT DATA SCIENTIST / ANALYTICS ENGINEER

Data Analyst is a phenomenal entry-level role but gets written off as 'easy' or 'basic' despite requiring a certain level of expertise. The role of a 'Senior Analyst' involves a lot more experimentation and A/B testing knowledge that can partner incredibly effectively with Product Owners and Scrum teams. On the other hand, an 'Analytics Engineer' requires creativity, being design-oriented, ability to learn quickly and further execute.

RESEARCH SCIENTIST

The role of 'Research Scientist' is probably the first to be fleshed out and understood. Typically, this is attractive to people holding PhDs and is responsible for pushing the boundaries of AI in our society, primarily dealing with Deep Learning and Reinforcement Learning.

MACHINE LEARNING ENGINEER

A traditional 'Data Scientist' and 'Machine Learning Engineer' can coalesce to be called a 'Machine Learning Engineer (MLE)' and comprises building end-to-end machine learning systems. Today, most MLEs seem to be responsible for the post-production modelling and Data Scientists handle the realm before. This can be problematic over time with the change in responsibilities for long-term model deployments. Hence, it is advisable to have a team capable of end-to-end oversight.
As new tools are being introduced and Python/ Jupyter-Excel integration is being rolled out, the PoC work that gets done in Jupyter is likely to be work that Analysts (senior level probably, if it involves PoC for modelling) primarily do.

DATA ENGINEER

The role of a 'Data Engineer' is another one that's here to stay and is fairly well understood. Data Engineers curate and source datasets from existing data sources (such as lakes, warehouses, etc.) and have primary oversight on how the data streams into the modelling and deployment pipeline.

WHAT'S NEW

The imperative need for machine learning in the public sector

For a sector that is meant to assist constituents, the sheer volume of backlogs and delays in the public sector is alarming. The four-month wait for passports, relatively longer than the pre-pandemic average of 6 to 8 weeks, made headlines last summer. As other sectors have done recently, incorporating artificial intelligence (AI) and machine learning into routine government operations can offer the intelligence, agility, and edge required to expedite procedures and enable end-to-end automation of document-centric processes. Given the enormous number of existing legacy optical character recognition (OCR) and other automation systems, government agencies must understand that short-term solutions will not bring about significant change and long-term success.

Clearbuds: First wireless earbuds that clear up calls using deep learning

As we witnessed during the COVID-19 lockdown, meetings were organised online and many people got distracted due to background noise that interfered with crucial discussions.
ClearBuds distinguish themselves from conventional wireless earbuds in two essential aspects. First, they employ a two-microphone array wherein two synchronised audio streams produced by the microphones in each earbud give us information and enable us to distinguish sound precisely. Second, the compact neural network improves the speaker's voice even more. The speaker's voice is then isolated and enhanced further.

Machine learning is instrumental in treating depression

Machine Learning can be used to treat patients suffering from depression. ML helps in analysing various data points such as genetic variables, family history, medical history, and clinical history. In one of the first studies of its kind, machine learning strategy will be utilized to identify the best treatments for depression, particularly helpful for countries like India. If everything goes as predicted, this technology technique might also be applied in low and middle-income nations. Sangath, a 26-year-old Goa-based mental health research organisation with regional offices in Pune, Bhopal, and New Delhi, will work with AIIMS Bhopal on the US National Institute of Mental Health-funded study to take this initiative forward.

DATA SCIENCE AT WORK

SALIM AKHTAR KHAN

PGP-DSBA MAY 2021 BATCH

"Read and find out how Salim was able to identify the causes affecting the downtime and recommended possible solutions for increasing the server efficiency."

Salim Akhtar works as an Infrastructure Engineer for the past 10 years and has been handling the deployment, maintenance, and migration of IT infrastructure. His key responsibilities include maintenance of servers, networks, and storage devices, and monitoring the characteristics of the operating system in physical and cloud data centres. In addition to it, he troubleshoots the systems as and when required and collects data through event logs and other monitoring aspects. This data can be used to further analyze, troubleshoot issues, and even provide recommendations for development.

The problem statement, in this case, was that the uptime for the application, server and network was below the SLA (Service Level Agreement) than what was promised (in this case 99% uptime for infrastructure devices). This could further result in imposing a penalty on the organisation as the customer is not getting what was promised. The organisation further needs to think about how this can be fixed. The data from monitoring tools and device logs need to be utilised intelligently for analysis.

To counter this problem, Salim exported the data obtained from log devices using MS Excel. Thereafter, he applied Exploratory Data Analysis (EDA) and data cleansing for structuring the data. Once it was done, Salim carried out outlier treatment, data mining and variable transformation. This data was then used for building models wherein techniques such as
Random Forest, Machine learning and Logistic Regression was used. Further to this, the classification reports helped analyze the model accuracy.
Using the model built and analysis done, Salim was able to identify that there were hindrances in the performance of the application which further affected the downtime. It was observed that the reasons for server downtime being affected were primarily the low RAM of the machine (that caused application breakdown and battery malfunctioning) and the disruptions in the location where the server was installed (such as frequent power failure). In the case of network downtime, he concluded that the Internet Service Provider (ISP) was the major bottleneck in poor performance.

Once Salim identified the causes of failure, he recommended possible solutions. This featured increasing the RAM capabilities, battery replacement, and Backup ISP link to automatic failover which could improve the SLA considerably thereby reducing the risk of penalties. In addition to it, he recommended migrating infrastructure to cloud providers such as AWS and Azure. He felt that these platforms can remove the headache of infrastructure maintenance and are much more reliable (assure 99.9% SLA).

AI AT WORK

ASHISH KUMAR

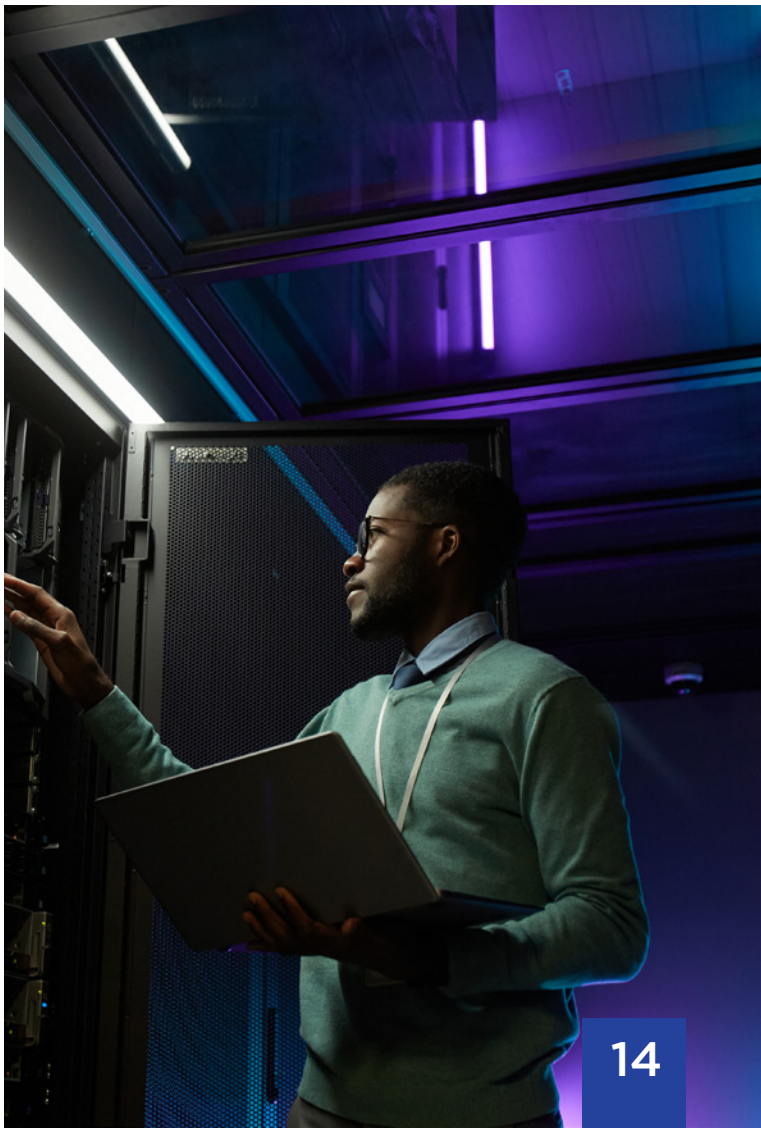
PGP-AIML ALUMNUS

"Read and find out how Ashish utilised AIML concepts to automate the process of printing and shipping in his organisation."

Ashish holds around 17 years of corporate experience. If we look at his last 10 year journey, he had been working with some of the biggest fintech companies. Currently, he is helping one of the top industry leaders in the credit card division and managing one of their legacy billing and statement platforms as a subject matter expert (SME).
He wanted to learn more about Artificial Intelligence and Machine Learning (AIML) to break the monotonicity and hence opted for Great Learning's PG program. He was fond of bioinformatics using AI but discontinued it due to salary constraints despite having a keen interest in this field. He read a lot of AIML to explore career opportunities and was
convinced that this is what he wanted to do.

He used the concepts learnt in one of the problem areas wherein their business had to share a 6-month advance inventory requirement with their print vendor. This was required for printing and shipping statements and letters to their 'Card Members'. Earlier, the process was executed manually and was quite erroneous. He identified this as an opportunity to apply some of his machine learning concepts and worked on automating it. He came up with a model capable of predicting by learning from the historical data set. The approach featured building a Time-Series Regression Model with an R2 score of more than 90%, which was much better than the manual predictive system being worked upon by a team of 5-6 members. Despite having a limited AI/ML opportunity available on his platform, he introduced a model that proved to be a breakthrough in the work process. This was a big success for the entire team.

He believes that by applying the concepts learnt in the AIML course he can make a successful transition toward a career in Artificial Intelligence and Machine Learning.



MASTER CLASS

REINFORCEMENT LEARNING

DR. DEEPAYAN CHAKRABARTI

ASSISTANT PROFESSOR,
UNIVERSITY OF TEXAS AT AUSTIN

Recently, we organised an industry session on 'Network Analytics' that was addressed by Dr. Deepayan Chakrabarti who serves as an Assistant Professor in the Information, Risk and Operations Management Department at the McCombs School of Business at the University of Texas at Austin.

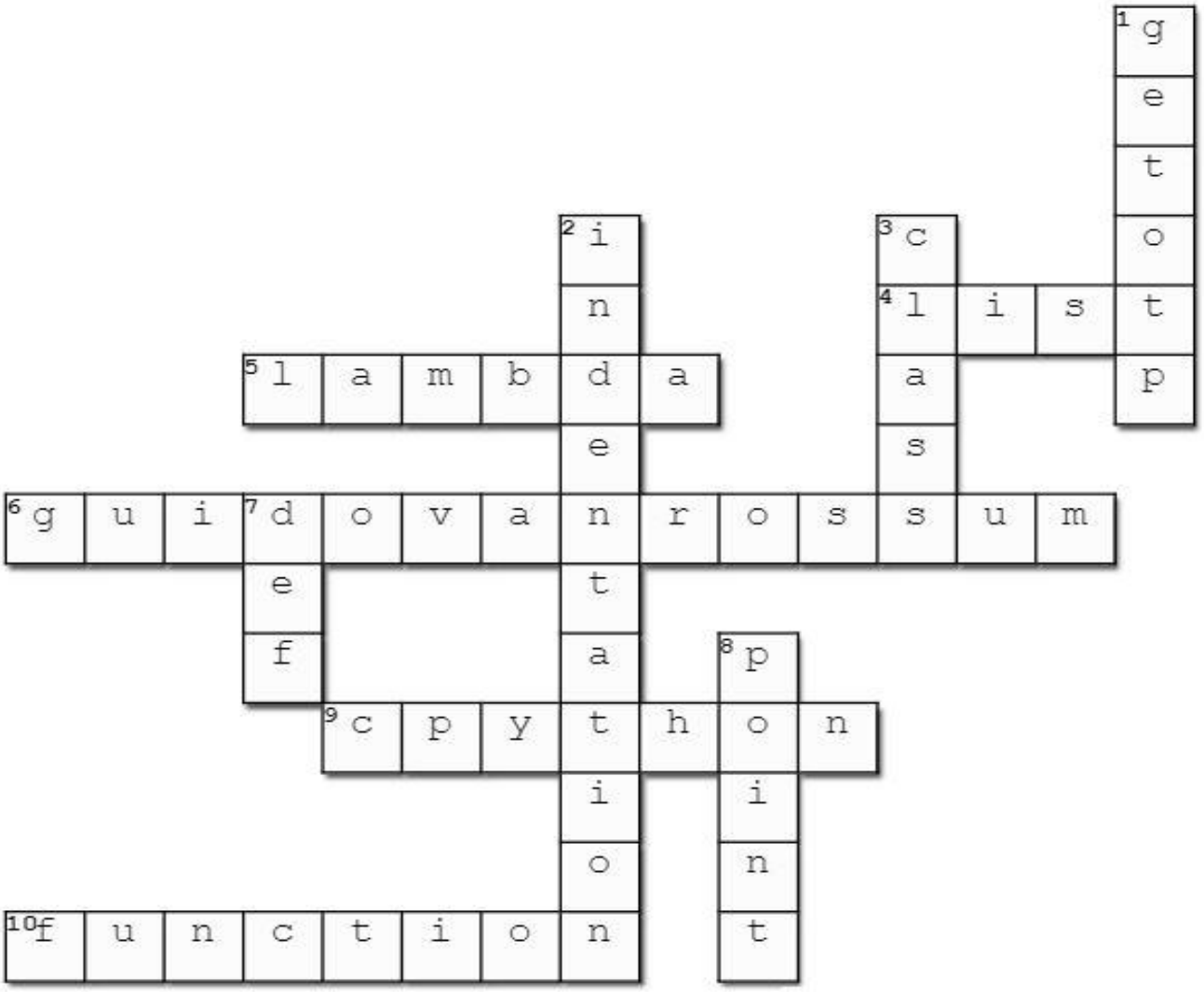
About Dr. Deepayan Chakrabarti -Dr. Deepayan Chakrabarti is a seasoned mentor and has completed his PhD in Computational and Statistical Learning from Carnegie Mellon University in 2005, and his B. Tech in Computer Science from IIT Kanpur in 2000. Subsequently, he worked at Yahoo! Research and Facebook, before joining the McCombs School of Business as an Assistant Professor at the University of Texas, Austin in 2014. He has authored about 40 peer-reviewed publications, 20 patents, 3 book chapters, and 1 book. He has rich expertise in solving problems related to Machine Learning and Data Mining, particularly focusing on mining large graphs and social networks, computational advertising, recommendation systems, and web search and information retrieval.

Session Highlights -
The session on 'Network Analytics' focused on problems faced by organisations with a large network of users (such as LinkedIn). He also stressed upon understanding the 'prediction problems' that such organisations encounter.

The webinar was useful in providing a broad overview of Network Analytics. The nature of the session ranged from basic to intermediate level. While one didn't need to have completed the ML course, some prior exposure to fundamentals would have helped in a better appreciation of this session.

The session focused on what is common between finding new connections on LinkedIn, recommending movies on Netflix, and tagging articles in a knowledge base? All of it was understood and analyzed in the context of a network. Indeed, these were just a few instances of prevailing fundamental problems in network analytics which one can call a 'link prediction' problem. The lecture gave a broad overview of methods that have been proposed to solve this problem. Some were simple and obvious, while others leverage ideas from deep learning. Yet, the same basic concepts lay underneath all of these methods.

CROSSWORD SOLUTIONS



ACROSS

4. Identify the data type from the code - L = [2, 54, ‘javatpoint’, 5]

5. Python supports the creation of anonymous functions at runtime, using a construct called?

6. Who developed Python Programming Language?

9. In which language is Python written?

10. Method used inside the class in Python language.

DOWN

1. Module in the python standard library parses options received from the command line.

2. Used to define a block of code in Python language.

3. Example of user-defined data type.

7. Keyword used for function in Python language.

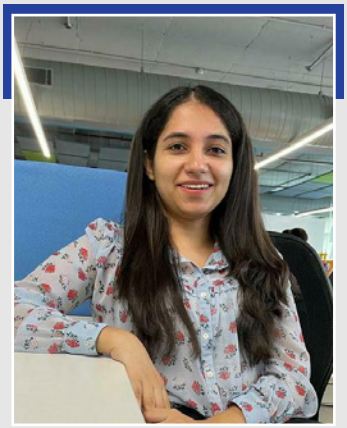
8. Output of the code - >>>’javatpoint’[5:]

LEARNING BIRD CHIRPS:

> "I realized that becoming a master of karate was not about learning 4,000 moves but about doing just a handful of moves 4,000 times."
>
> - Chet Holmes

THE EDITORIAL TEAM:

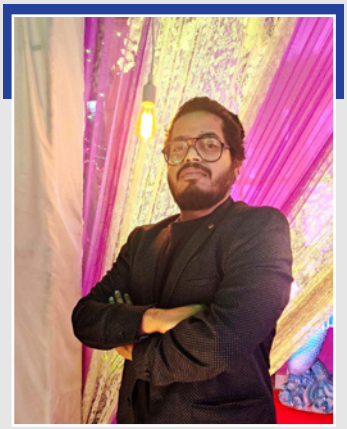
Surbhi Bhandari



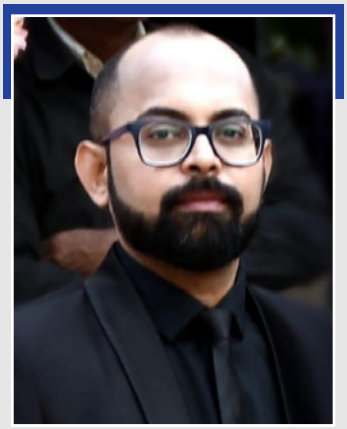
Nikita Duseja

Dipti Mayee Sahoo

Srijan Purang



Abhirup Dey



Mack Donald