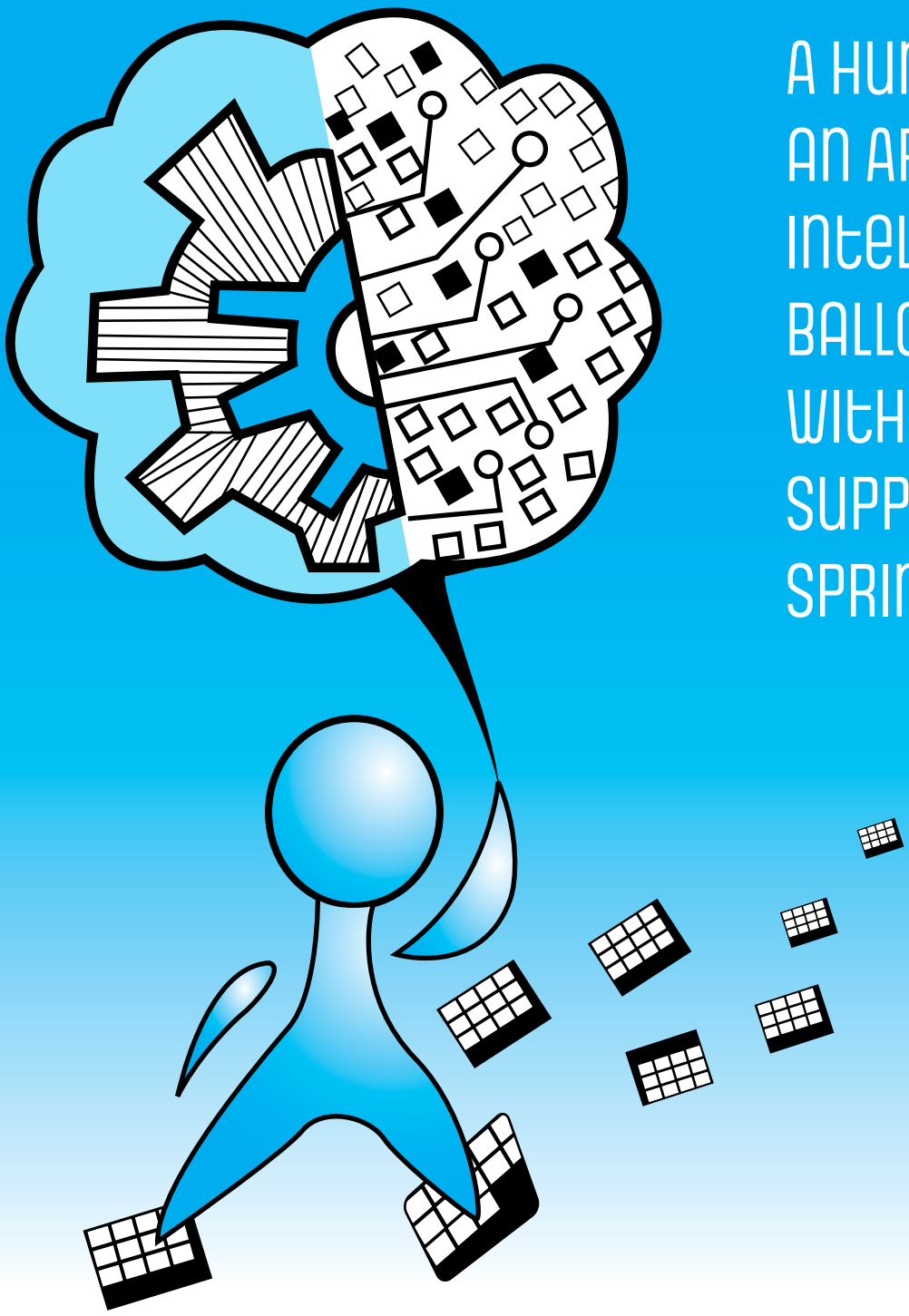
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY GURAJADA
VIZIANAGARAM COLLEGE OF ENGINEERING VIZIANAGARAM
DEPARTMENT OF INFORMATION TECHNOLOGY

NEWSLETTER

Bytes

September 2022 | Vol. 1 | Issue No. 1



A HUMAN CARRIES AN ARTIFICIAL INTELLIGENCE BALLOON WITH BYTES' SUPPORT AS A SPRINGBOARD.

Editorial

Dear Readers,

"To iterate is human, to recurse divine."

– L. Peter Deutsch

The Department of Information Technology is launching a newsletter, "Bytes," which encompasses state-of-the-art development in IT industries, competitive programming nuggets, Cutting edge problem-solving techniques, student and faculty achievements. This unique newsletter creates a platform for Alma matter to contribute articles on best industry practices. This approach helps current students to get acquainted with the latest technology. Active participation in alma matter is the primary goal of this newsletter. It serves the department and institution growth twofold. Firstly, the industry's best practices can be reflected in our student community through their alums. It makes more sense to the students. Secondly, it helps to bridge the gap between industry and academia. It's the dream of any computer science/ Information Technology graduate to join companies like Meta, Apple, Google, etc. The newsletter devotes a particular column to algorithm design techniques and their importance in MAANG companies. "Bytes" will act as a platform to nourish coding techniques to crack such companies. We invite our research collaborators and academicians to share a glimpse of fundamental research to enable our faculty and students to solidify their research quotient.

We are grateful to all of our contributors and readers. So many high-quality articles have been received. May God provide you with all the grit and well-being necessary to navigate these challenging times safely. We are indebted to the members of the Bytes publication committee, the editorial board, the authors, and the reviewers for their remarkable contribution and support in the production of this issue. Finally, I've traced and acknowledged ownership and copyright and obtained permission for items in this Newsletter. I will gladly contact any unreachable copyright proprietors to obtain rights. We would like to express our gratitude to our Hon'ble Vice Chancellor, Registrar, and Principal for their unwavering support, guidance, and encouragement throughout the publication of this issue. We respect the assistance and cooperation of our Information Technology department's whole personnel at the JNTU-GV University College of Engineering in Vizianagaram. We eagerly anticipate receiving constructive comments and recommendations from our respected members and readers at bytes.newsletter@gmail.com. **I'd like to specially thank Sails Software Solutions for sponsoring this issue.** I wish you and your loved ones a safe and enjoyable Navaratri and Diwali.

Dr. Tirimula Rao Benala
Chief Editor, Bytes - The newsletter

Editorial Board

What's Where

Message from Hon'ble Vice Chancellor

Dr. G.V.R. Prasada Raju

Hon'ble Vice-Chancellor *FAC*

"Learning is a continuous process from the moment we are born until our last breath."

I feel happy that the Department of Information Technology JNTU Gurajada Vizianagaram-College of Engineering Vizianagaram is launching a departmental Newsletter, "Bytes," which is long awaited. As you know, the holistic development of the Institution's students goes beyond classroom teaching. Thus, any such attempt to offer multi-pronged opportunities to its students and further help them to grow and realize their true potential is always welcome. This Magazine provides a platform for students for their creativity, self-expression, learning experiences, and even an identity they will be proud of through accepting the challenge of writing. Such an opportunity assumes importance in today's age of competition, in which the young generation seems caught up by the much attractive and addictive social media and networking. However, this cannot be a substitute for the human mind engrossed in thinking, exploring, and constantly introspecting about the self, surroundings, and country. This magazine allows its readers to delve into and appreciate new voices. All contributors and alums deserve congratulations. I hope other students will also feel inspired and motivated to improve their writing and presenting skills. I look forward to all students reaching new heights in their respective fields and making our College a proud institution. I congratulate the entire Editorial team for bringing together this edition of the magazine. Best Wishes, Jai Hind.

Message from Registrar

Dr. Swami Naidu Gurugubelli

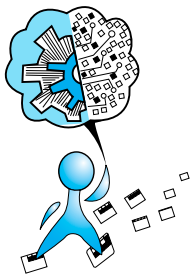
Registrar, $_{i/c}$
M.Tech (IITM); Ph.D
Professor in Metallurgical Engineering

It gives me great pleasure to know that 'Bytes' Departmental Newsletter 2022 is ready for publication. Providing ample opportunities in engineering education is one of the most fundamental obligations we owe to our students because, in JNTU Gurajada Vizianagaram, we are driven by the belief that every student deserves a high-quality education. 'Bytes' provides an intersection of significant challenges and an excellent opportunity for the students to review their efforts and analyze their research and development achievements. Technology is evolving at a dizzying rate, and our classrooms may not be designed to keep pace with it. There may be a lot wrong with the style of education. Still, the pages of 'Bytes' tell the tale of all that have been a part of what is right about the education they get in the Department of Information Technology, JNTU Gurajada Vizianagaram-College of Engineering Vizianagaram. I congratulate the Head of the department, a team of students, and the faculty for their tireless efforts that have come to fruition in this magazine. I wish it all success and hope the following generation of students will carry through this tradition that the current students have set. True to its name, this magazine gives an insight into the range and scope of the imagination and creativity of our students and faculty members. I applaud the editorial team for their hard work and dedication in realizing this goal, and I wish my dear students success in all future endeavors.

Message from Principal

Prof. K. Srikumar
B.Tech., M.Tech., Ph.D.
PROFESSOR & i/c PRINCIPAL
JNTU-GV University College of Engineering

Nurturing creativity and inspiring innovation are critical elements of a successful education, and a Departmental magazine is the perfect amalgamation of both. It harnesses the creative energies of the academic community and distills the essence of their inspired imagination in the most brilliant way possible. Hence, I am delighted that the Department of Information Technology is launching a departmental magazine, "Bytes, " ready for publication. I take this opportunity to congratulate the editorial board for bringing out this magazine as per schedule, which is an achievement considering the effort and time required. May all our students soar high in uncharted skies and bring glory to the world and their profession with the wings of education!



Algorithmic Bias and Multi-label Classification: Topics of Importance in Machine Learning and Data Science

Satchidananda Dehuri, SMIEEE
Department of Computer Science, Fakir Mohan University, Vyasa Vihar, Balasore-756019, Odisha, India.
Email: satchi.lapa@gmail.com

Machine learning (ML) [1] and data science (DS) are two important dominating fields in recent days in the area of computer science and information technology. They are increasingly used as effective and efficient tools for solving all most all problems occurring in every sphere of life. The dramatic growth of algorithmic decision making under the umbrella of machine learning and decision science continues to gain momentum in marketing, research in this stream is still inadequate despite the devastating, asymmetric and oppressive impacts of algorithmic bias on various customer groups. There are advantages to algorithmic decision-making; unlike human, ML and DS agents does not become tired. However, like human, ML and DS agents are vulnerable to discrimination that renders its decisions ''unfair". Discrimination is the unfair treatment of individuals based on specific characteristics, also called sensitive features such as gender and race. It has been found that machine learning not only leads to unexpected results with bias, but also has amplified algorithmic bias. Extensive studies have been conducted to achieve the fairness in ML and DS model, and generally categorized in three approaches: preprocessing, in-processing and post-processing. Pre-processing solves the problem by eliminating the bias present in the training data itself. In-processing reduces the bias by adding a constraint to the learning algorithm even if there is a bias in the data. Postprocessing ensues decisions themselves. Readers may refer [2] for more details on algorithmic bias. Now we will turn our attention to second part of our topic of discussion i.e., multilabel classification.

Multilabel classification, an equally important topic of ML and DS aims to build classification model for instances assigned with multiple-labels simultaneously which is a common learning paradigm in real world application [3]. It deals with multiple labels being assigned to every instance in a dataset which can be assigned to more than one class simultaneously. Multi-label classification tasks exist in many real-world applications, such as, gene classification in bioinformatics[2], medical diagnosis, document classification, music annotation, image recognition, and so on. All these applications require effective and efficient multilabel classification algorithms. There exist a variety of multilabel classification algorithms[3]. The existing multilabel classification algorithms are developed based on two basic approaches like algorithm adaptation and problem transformation method. Problem transformation method is to transfer multi-label classifications into multiple traditional single label classifications, specifically, multiple binary classifications. After a multilabel classification problem is transferred into multiple binary classification problem then, all the traditional classification algorithms can be applied directly to build a classifier for each binary dataset and make prediction for its correlated test instances. For more details about multilable classification, readers may refer a tutorial authored by Gibaja and Ventura [3].

References:

[1] T. G. Dietterich. "Machine learning," in Encyclopedia of Cognitive Science, vol. II, L. Nadel, Ed. London: Nature Publishing Group, 2003, pp. 971–981.
[2] A. Koene, Algorithmic Bias: Addressing Growing Concerns, IEEE Technol. Soc. Mag. 26 (2) (2017) 31–32.
[3] E. Gibaja and S. Ventura, "A tutorial on multilabel learning," ACM Computing Surveys (CSUR), vol. 47, no 3, pp. 52:1–52:38, April 2015.

About the Author

Satchidananda Dehuri, TARE Fellow is working as a Professor in the Department of Computer Science, Fakir Mohan University, Balasore, Odisha, India since 2013. He received his M.Tech. and Ph.D. degrees in Computer Science from Utkal University, Vani Vihar, Odisha in 2001 and 2006, respectively. He visited as a BOYSCAST Fellow to the Soft Computing Laboratory, Yonsei University, Seoul, South Korea under the BOYSCAST Fellowship Program of DST, Govt. of India in 2008. His h-index as per Google Scholar is more than 25. He has already published about 200 research papers in reputed journals and referred conferences.

Security Risks of Suspicious Data reduced by a Novel Model

Gullipalli Suneetha

NIT Calicut M.Tech (CSIS), Alumni of JNTUV(IT). Email: suneetha21095@gmail.com

Introduction:

Modern operating systems are notable for being open, allowing us to download data or run programmes from any Internet source, reputable or not, without hesitation. When we wish to use this data or codes and maintain the system safe at the same time, a paradox arises.

Even after decades of research and experience, this issue is still seen as a major obstacle. In order to lessen the security concerns brought on by this dubious data or programmes for open operating systems, this study provides a revolutionary dynamic defensive model (DDM). The elements enable DDM to provide complete, dynamic, and real-time security protection throughout the operating system's entire life cycle.

Dynamic Defense Model:

A high-level security defense abstraction with four essential parts is called the "Dynamic Defense Model": 1. Marking labels dynamically, 2. Tracking labels dynamically, 3. Modulating labels dynamically, 4. Run-time regulation.

First of all, we cannot completely trust them given all the questionable data and codes. As a result, we must give them a distinctive label to set them apart from reliable information or codes.

Second, when the OS is functioning, suspect and sensitive data are consumed and spread across the system, and suspect processes produce child processes. We need to monitor where this data flow, how much data was contaminated, and how these shady activities impact the system via dynamic label tracking.

Thirdly, the original label may not adequately describe the suspicious processes' current security station after real-time execution of the suspicious processes was seen. According to the suspect processes' most recent activity, the label should be dynamically modified.

Last but not least, we shouldn't let any data or programmes spread or operate freely in the system without any controls.

Data Label Marking:

DDM typically offers three different types of labels: risk, cap, and sens.

Risk label: As soon as they enter the system, we should identify all the suspect data andcodes that originate from Internet sources with specific names so that we can distinguish them from the reliable data and codes.

Cap label: The cap label is intended to stand for the most fundamental or minimal capabilities (permissions or privileges) that the suspicious code should be granted in order to carry out its intended function.

Sens label: The sens label is designed to represent how important the sensitive data or resources are.

Label tracking:

We define the high level rule of our label tracking as:

- For the risk label, this rule can be specified as:

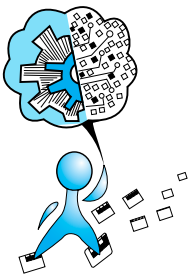
  $A \sim B \Rightarrow B.risk = max(A.risk, B.risk)$

  Label tracking is the same for both risk and sens labels.
- For the cap_o label, this rule can be specified as:

  $A \sim B \Rightarrow B.cap\_o = (A.cap\_o \cup B.cap\_o)$

Dynamic Label Modulating:

DDM modifies the label in three different ways: automatically incrementally, automatically decrementally, and Evaluative modulation.



The rule for automatic incremental modulation of the risk label can be described as:risk_new = risk_old + f (action_ill)

Similarly, the rule for automatic incremental modulation of sens and cap_o label can be described as: sens_new = sens_old + f (action_sen)

cap_o_new = cap_o_old + f (action_sen)

The rule for evaluative modulation can be described as: label_new = f (actions)

Runtime Controlling:

We cannot let any data or codes to spread or operate at will in the system without any controls. To avoid any potential harm to the system or the loss of confidential information, we should stop harmful acts in accordance with security regulations.

The following are examples of high-level runtime regulating rules: For the risk label, this rule can be specified as: running is permitted ⇐⇒ risk < threshold_kill

According to this rule, a process can continue to run if and only if the value of its risk label is lower than the process-killing threshold. For the risk label and sens label, this rule can be specified as: access is permitted ⇐⇒ risk + sens < threshold_sys

According to this rule, a process with a risk label can only access resources or sensitive data if the sum of the risk and sens values is below the system threshold. For the cap_s and the cap_o label, this rule can be specified as: access is permitted ⇐⇒ cap_o ⊆ cap_s

According to this rule, a process with the label cap s can only access resources or sensitive data with the label cap_o if and only if cap_o's value is a subset of cap_s.

Conclusion:

When we wish to employ dubious data or code from an Internet source that we can't completely trust while maintaining the operating system's security, it might be quite difficult. A unique dynamic defensive model is presented by the author as a solution to this issue, reducing the security threats posed by suspicious data or codes.

Reference:

[1] Reducing Security Risks of Suspicious Data and Codes Through a Novel Dynamic Defense Model Zezhi Wu, Xingyuan Chen, Zhi Yang and Xuehui Du.

About the Author

Gullipalli Suneetha is currently pursuing her Masters in the stream of Computer Science and Engineering with a specialisation in Information Security at the National Institute of Technology Calicut. Her interest lies in data and information security. Also, she is working on UAVs and their security, as drone and related attacks are becoming more dangerous nowadays.

Let's face it.

Ch. Harshini

Software Engineer in Microsoft. Email: harshini.chekuri97@gmail.com

Landing the job you want isn't always a straightforward process. Everything from spellbinding your resume to exhilarating your interviewers can be challenging and completely nerve-racking. Top companies not only look for people who work hard but people with a positive attitude. Always come up with a great story that speaks to your dedication, hard work and perseverance. And you've heard it before, because it's true: "**Communication is key**", always try to be proactive and express your interest, this mantra will definitely place you in great hands.

Being a fresher, you need to get your foot in the door. A generic resume isn't likely going to do it, no matter how polished it looks. So make sure it lines up with the job description, because any communication to the hiring managers should be very clear in a way you're a fit for the role.

Finally, are you juggling around the interview process? Here is the life saver: "**Always make coding your priority**". There are many platforms available to enhance your coding skills like Leetcode, HackerRank etc. You can't just go with the last few days of preparation to companies like Google, Amazon and Microsoft. Your learning should be a continuous process. The better you are with your core knowledge, the best you will get.

Speaking out your thoughts to interviewers makes a big difference. Sometimes you might not solve the problem as expected but might end up with a job. Sounds weird right? But it's true. These days all top companies are looking not only for problem solving skills, but also expect you to communicate your thoughts. Remember, they mainly focus on your approach to the problem which can be achieved only by putting your thoughts out.

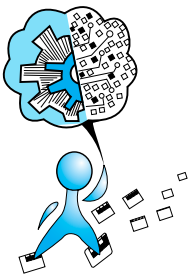
Your best bet to improve your skills is collaborative programming using shared environments in real-time with your friend or colleague by sharing each other's thoughts and solving problems together. Don't be discouraged if you can't solve it in the first go. Keep trying, the more the merrier. It always takes hard work to achieve something important.

"**Don't be afraid to fail**". Gaining experience is more important. The more interviews you give, the more confidence you gain back.

LinkedIn profile: https://www.linkedin.com/in/harshini-chekuri/

About the Author

Ch.Harshini, currently working as a Software Engineer in Microsoft. She did her Masters in Computer Science at West Virginia University, USA.



Data Pipelines architecture and basic understanding

Lohit Ravi Teja Bhupati

Data Engineer, Walmart. Bentonville, AR. Email: lohit.raviteja008@gmail.com

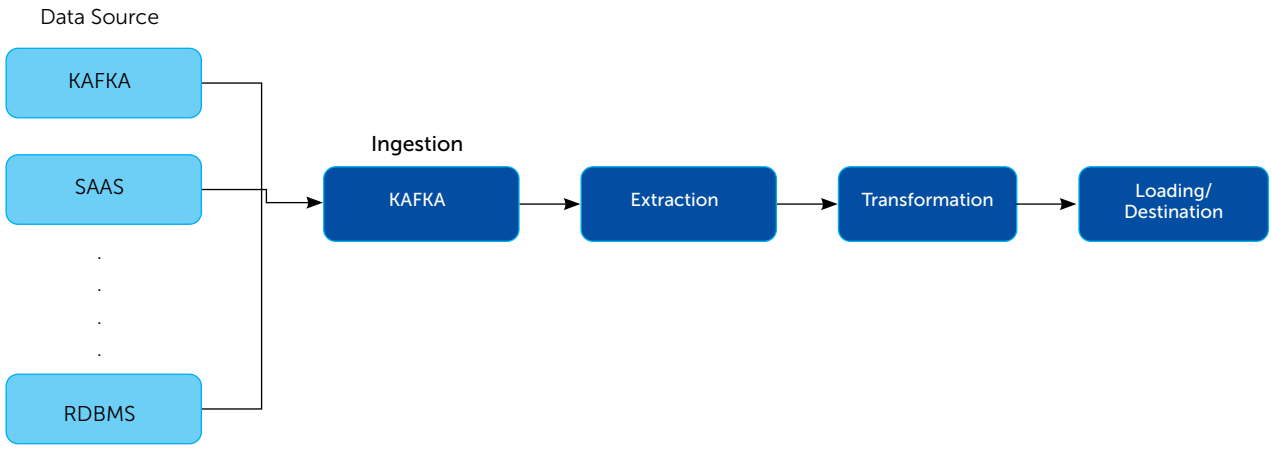
Data pipelines are the series of steps performed to process the data into required formats/ destination sources. In these pipelines the output of each step will be source to the following steps. There are different scenarios where the pipelines are designed, it can be a simple batch job to read raw data and convert it into formatted data and saving into sources tables, consuming data from the upstream applications and perform transformations so that the data could be utilized by the downstream applications or creating visualizations using the transformed data etc.

There are three major factors that contribute to the speed with which the data flows in a pipeline, **Throughput** can be defined as the rate at which the data flows through the pipeline. **Latency** is defined as the time required for a single unit/token of data to flow through the pipeline.

Reliability is the measure completeness and accuracy. There are different techniques that are in place to ensure the data reliability like logging, auditing, validating etc.

There are some basic questions one should have before building a pipeline.

- What data source?
- What type of ingestion?
- How frequent should the load happen?
- How should the end data look like?
- Is the data consumed by any downstream applications or any visualizations are to be built on top of it?



Data Pipeline architecture

Data Source:

It can be considered as the first and key layer of the design. It may include the data from any streaming sources, SaaS applications and relational databases etc.

Ingestion:

It is defined as a process to read the data from the data sources. We can read the data from each source using the API's provided by the data sources itself. Before reading the data, we need to analyze the data sets to gain insight

Contd. on page 18...

About Hi-Track

BalaSai Charan Mamidi

B.Tech from JNTUK, India. Email: mamidisaicharan@gmail.com

Hemp Industry Track or HI-TRACK is an integrated traceability platform specifically built for Hemp industry in Australia, developed by Natsoft Consulting in association with HempGro to address the challenges faced by Hemp growers and regulator in WA.

It is designed to bring all the stakeholders on board on common platform to help regulation with data and compliance processes. Digital technologies such as Blockchain, IoT and Machine learning enable much more possibilities and superior data security.

It has custom modules for Farmers, DPIRD (regulator) and Seed Testing Agency. Web and Mobile Apps to input and view data at each stage of the crop from seed to offtake process. It provides birds eye view of movement of seed stocks across farms to the regulator to enable.

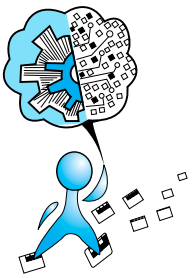
Benefits:

- Ability to deal with compliances easier with provenance of activities from seed to offtake recorded on Blockchain.
- Continuous record of vital parameters through out the growing and processing period which can help repeat good crop performance

- Ability to provide historic data for compliance, planning and automating filling of renewals for every new crop
- Ability to have geo-tagged photos of crops at various stages for records and to share with regulator along with
- notifications for sowing, flowering and harvest.
- Ability to have unique identifiers for incoming seeds and outgoing produce, in the form of QR code

About the Author

BalaSai Charan Mamidi is a Blockchain Solution Architect with over .5 years of experience in designing and development of blockchain solutions. He has built several blockchain solutions in Supply chain, Agriculture, Land Registry and Healthcare domains. He also holds a patent for 4 layered IoT Architecture for Hyperledger Fabric Blockchain based Self-managed vehicle ad-hoc networks, granted by IP Australia.



Data Quality and Machine Learning: What's the Relationship?

Jitendra Yasaswi K.

B.Tech IT, 2014. Research Scientist at the Robert Bosch Research and Technology Center, Bangalore.
Email: katta.jitendrayasaswibharadwaj@in.bosch.com

The buzzword AI is an increasingly used term in today's technological world. In the last decade, we have witnessed enormous growth in AI enabled applications like never before. At the start of last decade, the focus of Machine Learning (ML) applications was to tag or classify images and today we have ML algorithms that can even generate images (e.g. DALL-E 2). Similarly, a decade ago we started with applications that can predict sentiment of a user from movie reviews and today we have systems that can generate text and even fill parts of software code (e.g. Github Copilot).

Many ML practitioners consider the year 2012 as the breakthrough year for AI where they started to believe that deep learning models indeed work better than traditional ML methods. This success can be attributed to two important factors 1) the availability of large training datasets and 2) the availability of infrastructure to train large deep learning models. If you look at any ML problem setting in the real world, it has two main parts which are the *core model* part and the *core data* part. The core model part deals with designing models (say deep learning models), optimizers, coming up with suitable training schemes, validating and maintaining these trained models. As a result of this, we have witnessed a number of important technical developments on the *core model* part like improvements in Convolutional Neural Networks (CNNs), Variational AutoEncoders (VAEs), Generative Adversarial Networks (GANs) and Transformer Models. On the other hand, talking about the *core data* aspect, apart from collecting huge datasets, very little or no emphasis is laid on the quality of the training data and managing it.

Many challenges that AI based applications face in the real world are due to imperfections in data. Imagine a scenario where you are developing an AI algorithm which is to be deployed in a real world application. To ensure that your AI algorithm works with extreme reliability, this algorithm has to be trained with data that contains even rarest of the rare situations that the application might encounter in the real world. Recently, Tesla's autopilot feature had mistaken a horse carriage for a truck and this could be due to the long tailedness of the data (the ML model(s) deployed in autopilot may have encountered very few or no scenes containing horse carriage during training). This is why "*Data-centric AI*" is needed. Here, the focus is on collecting good quality training data and ensuring good quality labelling. This also includes ensuring good data coverage, which could be done through search, retrieval of rare instances, usage of data augmentations (self-supervision) and synthetic data generation to generate rare scenarios or to fill domain gaps. This also includes methods to identify and handle data drifts post model deployment.

There is a saying that "*Your model is as good as your data*" and good quality data is a must to build any successful ML pipeline. As articulated by Andrew Ng in "MLOps: From Model-centric to Data-centric AI", the future is moving towards Data-centric AI.

About the Author

Jitendra Yasaswi Katta is a Research Scientist at the Robert Bosch Research and Technology Center in Bangalore, India. His work spans the intersection of Computer Vision and Machine Learning, with special focus on self-supervised learning and representation learning. Prior to this, he worked at Teradata Labs where he was part of the R&D team developing Machine Learning applications for prediction tasks. He received his Master's degree (MS by Research) in Computer Science and Engineering from IIIT Hyderabad and Bachelor's degree in Information Technology from JNTUK UCE Vizianagaram.

Entrepreneurship – How I launched my new Startup

Pratap Chillakanti,

PhD, MBA. Email: pratap@rewardscloud.com

India is now the fifth largest economy in the world. One of the significant reasons for India's resurgence is the entrepreneurship revolution that is sweeping the country. Numerous factors contributing to this revolution include 1) the ground breaking work done by the large information technology companies in providing product development services for the innovative multi-national companies, 2) peoples exposure and awareness of the technology innovations that are possible, 3) enthusiasm of the new generation of Indian students and workforce that aim big, 4) national and state government policies that encourage and support the launch of startups, and 5) generational inspiration from the fact that Indian-born leaders are now the CEOs of marquee companies like Google and Microsoft.

For aspiring entrepreneurs, once the thought occurs about launching their own startup, they do not necessarily have a clear roadmap to manifest their thoughts into an action plan. In this paper, I will narrate my experience over the last two years that allowed my startup to take birth, and progress toward a product launch in the summer of 2022 and landing a marquee client Google.

1. Background

In 2014, one of the iconic Silicon Valley networking companies hired my company to assist them with one of their channel partner incentive and rewards program. Numerous companies sell their products to end customers in two modes: 1) direct sellers – meaning that a company a has their own sales teams who sell the products and solutions to the end customers, and 2) indirect sales – selling through intermediaries, typically called channel partners. These channel partners interface with the end customers. The channel partners typically sell products and solutions from different companies. To ensure that a partner positions their products to a customer, the companies design a variety of incentives and rewards so that a partner is more engaged with them and that they position their product/solution to the end customer.

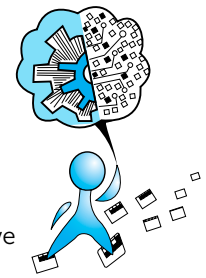
In the case of the incentives program that my company was program managing, we were interfacing with a vendor who provided the software solution that managed the incentives and rewards for the individual partner sellers of a partner company. Over a course of several years, there was partner dissatisfaction on several fronts:

1. The process of submitting a purchase order (PO) was error prone and the system could not find the PO because of syntax errors. As a result, they were not getting the rewards on time.
2. Partners opened support cases to resolve such issues. Since many POs were not processed automatically, the support cases began to grow rapidly, the resolutions times increased, and the partner dissatisfaction increased.
3. The technology and the software algorithms in essence were outdated. Some of the processes were manual and/or semi-automatic.
4. The company management was frustrated that partners were expressing their dissatisfaction, their engagement levels dropped, and the Return on Investment (ROI) was decreasing.

There was an opportunity for innovation, automation, and transformation. It was time to think about a new approach to solving these challenges. As an entrepreneur, I sensed an opportunity to solve a customer problem. We proposed a new state-of-art solution to address these challenges, enhance the system functionality, automate the process, and ensure a high level of partner satisfaction. Once approved, our team of almost twelve people designed, developed, and deployed a solution within nine months. The system went live in the first quarter of 2019. Since then, our team supports an ecosystem of over 25,000 partners.

2. The Beginning

At the beginning of 2021, I decided to launch a new startup, Rewards cloud Inc (https://rewardscloud.com).



The purpose of the startup is to develop a solution in the channel incentives and management to begin with. My team and I had acquired domain knowledge including a deep understanding of the still unresolved problems and potential for innovation and transformation. To commit to this new journey, we needed to first do due diligence about its feasibility. Afterall, it requires a commitment of time, money, and resources. The following is a high-level process that was followed to arrive at a decision.

1. Collaborative research on channel partner ecosystem

First, I gathered an excellent team of trusted advisors who worked in the industry in the channel partner space. These advisors were pioneers and thought leaders who have defined and led numerous channel partner programs in their distinguished career. The team also included my technical leads who gained expertise because of the development of similar solutions. We took about one month to do focused research to understand the channel partner ecosystem and the various sub-categories embedded in it. We had a recurring meeting, and we had a shared folder to share our work in progress and our observations.

2. Collaborative research on incentives and rewards

A team was dedicated to researching the incentives and rewards landscape beyond the channel partner ecosystem. For example, there are many loyalty programs in industries such as air travel, hotels, and credit card programs. The purpose was to identify motivation factors that trigger people to act the way they do, and how we could use the theory of motivation to influence certain behavioral outcomes. This aspect of the research also lasted about a month.

3. Scoping the problem and the opportunity

The team evaluated the results of steps 1 and 2 continuously. We identified numerous directions that we can take to develop the solution. For every direction suggested, we did a further evaluation of the product and solutions that are already in the marketplace. After an iterative process that lasted almost a month, we finalized the scope of the problem and the opportunity.

4. Competitive Analysis

Now came the critical task of doing a comprehensive competitive analysis. For an entrepreneur, this is a crucial step in the journey. There are several aspects to conducting a competitive analysis. The product that is envisioned must address an either an "unmet" need in industry, or perhaps identify a new and innovative way of solving existing problems. Once the scope was defined, the team dedicated itself to identifying about fifteen different vendors in the marketplace that address problems like what we were trying to solve in an innovative fashion.

5. Identifying the differentiators

This is yet another task that requires a deliberate and objective approach to ensure that we are embarking on the right journey. The differentiators must distinguish the startup from other competitors in such a way that once the product/solution is released the barrier to entry is set very high. It means that it will take a long time for competition to catch up. The team identified five differentiators, but we needed validate whether these were going to be true differentiators.

6. Expert evaluation of the differentiators

Through the contacts of our industry recognized advisors and our clients, we conducted several interviews to get valuable input on these differentiators. The experts helped us evaluate these differentiators multiple perspectives. The result was a more holistic way of combining these differentiators to create a powerful Partner Ecosystem platform that is beyond just the incentives and rewards management.

7. Making the Go/No Go decision

This was the last step of the process. The entire team met for a final discussion where all the research, the methodology adopted, and the final conclusions were presented. After some passionate discussions, the team unanimously approved the decision to go to the next step of identifying the Minimum Viable Product (MVP) that needs to be developed.

3. The MVP Development Journey

The scoping of the problem and the opportunity is at a broader level as what we aim to achieve with our product/solution. In the dynamic product introductions in today's world, it is important that startup products be introduced into the market quickly to ensure their competitive advantages. This implies that in general the product development lifecycle should be such that the new product is introduced in a reasonable timeframe. In our case, we determined and agreed that we should introduce our product within a year from the product development launch date. The business and the technical teams collaborated to first document a detailed list of the product requirements over a course of two weeks. The technical team further evaluated these requirements in terms of time estimates for completing those tasks. Through an iterative process we all agreed on the subset of the overall requirements that will be part of the first release of the product. It is a sufficient set of requirements encompassing the five key differentiators. This became our MVP.

The MVP determined the team structure including the knowledge, skills, and experience required for different roles such as technical architect, UI/UX designer, front-end developer, back-end developer, DevOps engineer, Web Developer, Testing and QA lead, and technical project manager. Since our startup did not have all the required skills in house, we attempted to hire the right people but were not able to put together a good team on time. A strategic business decision was made to hire a technical consulting firm to provide us the necessary resource. The development started in earnest in the first week of June 2021.

4. The Challenges and The Solutions

The development process had its own share of challenges that required very active people and process management. Some of the challenges that we faced include the following:

1. It was clear very early on that some of the back-end developers provided by the consulting company were not competent enough to meet our stringent standards of algorithm design and coding. They had to be replaced with substitutes which then resulted in a few weeks of delays in the development process.
2. Some of the development team members provided by the consulting team company were not their direct employees. We noticed that their commitment levels and their motivation was not on par with the rest of team. As the overall leader of the team, I worked with the tech leads to device ways to motivate the team members go beyond the normal call of duty to achieve the development milestones in a timely manner.
3. There was a crucial product demonstration meeting a highly influential analyst who agreed to review our product and mention it in his widely circulated report about the state-of-art I the channel partner domain. To meet this deadline, we had to overcome the above-mentioned challenges and inspire the team to step up to the challenge. This entailed working through the weekends and doing whatever it takes to succeed. We passed the analyst gate and eventually our product was included in the analyst's report.
4. At the end of February 2022, we had a rude shock from the consulting company that the two critical back-end resources that the company hired through third parties will quit the project with one day notice. This essentially put the entire project in a jeopardy. Such delays would cost a startup precious time and significant money. Fortunately, we were able to solve the problem by hiring two better backend developers within a week and the development was back on track.
5. The MVP was finally delivered at the end of May 2022 as we estimated. In the following three months, the teams focused on enhancing the security of the product so that it is compliant with the various Infosec standards in the industry.

5. Landing a Marquee Client

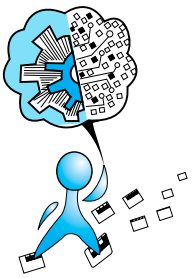
During the journey, the team gave numerous demos through a combination of prototype/interactive mockups to several companies. This allowed us to get valuable feedback from prospective clients. One client, Google, saw our demo and got insights into how we developed the solution for the networking client. They signed us up to implement their customized solution. Our team got onboarded and withing six months, we hope to deploy the custom solution.

6. Conclusions

The journey continues. The team finalized the requirements for the next phase of the product. This article aims to shed light on what it takes for an entrepreneur to bring a product to life once they get a small idea and they start dreaming big about the changing the world. The article did not touch upon the other important aspects of such a journey. Most importantly, it does not present the approach to financing the product development. In our case, we had a paying client that generated positive cash flow. In addition, as a founder of the company, I invested my personal funds into this effort. For more details about how to finance entrepreneurial efforts, you can contact the author.

About the Author

Dr. Pratap Chillakanti holds two technical PhDs, one in Information Security, and one in Transdisciplinary Research and Education. He also holds an MBA from University of Chicago Booth School of business. Pratap is an Entrepreneur, Adjunct Professor, Researcher, and a Mentor. He has over 30 years of experience in both the Silicon Valley technology industry and the academia



What technology is driving Zepto's 10-minute grocery delivery service?

Mediboina Sai Kumar

M.Tech '23 @ NIT Surathkal, Karnataka. Email: msaikumar6789@gmail.com

The digital transformation of industries has been fairly dramatic, and the grocery industry is no exception. When someone said, "Let's go shopping," it used to mean going from store to store in the neighbourhood, which took a lot of time. However, the required information may now be verified quickly, and once more, you can make a payment even if you have no cash on hand. This is how the process of digitization has advanced us significantly. Everything around us is being taken over by online enterprises. Due to their rapid growth, users' purchasing experiences now place a premium on speed and efficiency. Online commerce has changed the way that business is now conducted has changed, and customers can now relax in their homes and order anything from anywhere knowing that it will arrive on time, at a reasonable price, and at their doorsteps.

Have you ever heard of a grocery delivery app that takes just ten minutes?

You did read that correctly. After you place your order, groceries will be delivered in just 10 minutes to your door! Introducing the Zepto app, the quickest grocery delivery service in India.

Zepto, grocery delivery within 10 mins:

Aadit Palicha and Kaivalya Vohra, two 19-year-old Stanford dropouts who created the 10-minute fast grocery app Zepto. It is a Mumbai-based firm that provides 10-minute deliveries of necessities. It was launched in 2020. Customers have access to thousands of well-liked items, including dairy, groceries, farm-fresh fruits and vegetables, and more. According to the business, an average delivery takes about 8 minutes and 40 seconds. The 10-minute food delivery service employs an inventory approach to operate, delivering groceries from nearby micro-warehouses.

Zepto created dark stores as a novel solution to some of the execution issues that other supermarket applications have been experiencing. These shadowy warehouses have enabled Zepto to streamline its processes and provide its customers quick deliveries. Dark storefronts have allowed the firm to expand further into locations with heavy traffic, which has improved those cities' conversion and retention rates.

Tech behind this?

The idea and implementation of Dark Stores have been crucial to Zepto's success. So, what actually are these dark stores and what obstacles did they help overcome?

Retail grocery stores that exist solely to fulfil last-mile pickup and delivery orders are referred to as "dark stores." They are often found a few miles from the consumer, are smaller than a conventional grocery shop, and have a lesser selection of goods in exchange for delivery timeframes of less than an hour.

The idea is not new, and similar stores are widespread in the USA, Russia, and other developed economies. However, the difficulty in beginning operations in India was due to the nation's chaotic traffic and crowded streets. As a result, the location and planning of the "Dark Stores" became crucial. Zepto appears to have perfected the methodology used to choose these outlets' locations in India, since it presently has more than 40 dark stores in various cities.

Connectivity is the second issue that the dark stores have

shown to help Zepto with. Without dark stores, last-mile connection would not have been possible due to the breadth of the country and the presence of people practically everywhere.

Managing the amount of supplies is another difficulty that India is accustomed to. Dark stores are normally used to stock a smaller range of goods, so Zepto must have given careful consideration to how to handle the variety that comes along with the volume of deliveries in each Indian metropolis.

Overall, Zepto has been successful with its setup and infrastructure approach and is on a solid growth trajectory.

Conclusion:

Zepto is making news because of how quickly it moves. Nevertheless, quick delivery does not sacrifice quality. One of India's start-ups with the quickest growth is this one. Customers are responding positively to the company's miraculous 10-minute product delivery. Their mission is to 'make 10-minute delivery normal'. Even though it seems so absurd to us, it's actually working...

About the Author

Saikumar Mediboina is currently pursuing his masters in Information Technology at the National Institute of Technology Karnataka Surathkal. His area of interest is networking. He is currently working in tactile internet to achieve the low end to end latency and high reliability Haptic media type with http and other protocols Synchronization of multi modal data distributed to multiple devices and location.

...Contd. from page 11

of the data through a process called Data Profiling, which is used to examine the quality and structure. The data can be ingested either batch ingestion or streaming ingestion. Batch Ingestion is a sequential processing where the set of data records can be extracted and processed together. These batch processes can be scheduled or triggered manually. In Stream ingestion a single record of data can be processed automatically from the data source as soon as it is created or in time windows which can be used to give near real time data.

This ingestion or extraction frequency depends on the requirement on how the subsequent application need the data to be loaded.

Transformation:

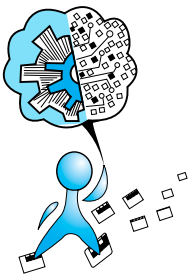
Once the data extraction is done, we can transform the data into the required format of the destination system, which helps to analyze the data.

Destination:

A destination is a data warehouse, a database or a data mart to hold the data after the data is processed.

About the Author

Lohit Ravi Teja Bhupati is currently working as Data Engineer at Walmart, Arkansas. He has done Masters thesis in Machine learning and Neuro imaging from University of Houston Clear lake, USA. With his interest towards Machine learning and data analysis, he is always upscaling by learning new technologies and continuing his passion.



Coding is not Everything A Software Student Does...

Motupalli Vijaya Sai Priyanka

Bachelor in Informtion Technology department from JNTUK-UCEV Vizianagaram. Email: mvspriyanka10@gmail.com

At the offset what everyone thinks that a software background student does is coding coding and coding. But believe me we aren't just coders. We never just code, its just a part of the journey. We do a lot more things other than coding. To make it simple let me pull out what coding actually is. We obviously need a language to communicate with one another. Based on the regions of the world, we choose our mode of communication. Its quite the same case with electronic gadgets as well. They need their own language to communicate as well. Just tell them we aren't coders we are language builders and the bridge between human world and the electronic world. We often listen to people saying we are nothing if these so called core students aren't there. What will we do if they didn't build a system..What if they did not assemble a robot? I just wanted to ask them, What will they do with their fully built computer system or an assembled robot if there is no proper language to communicate and make them understand what it is exactly intended to do? So, believe me we are never low for being coders, mark them we are never low!! We build a fresh language. Moreover we teach all the core built systems their cause of creation.

We often hear statements like coding is easy!!LOL!!. Not everything is easy and tough too. The perspective changes its complexity. As the era proceeds, the mindset starts changing for everything to be simple and fast. This probably can't work if software is not a part of it. Coding is just not the mix and match of English words with some special characters. It is possibly the way of turning complex into something simple and smooth. Let me remind you, the complexity starts with the builders themselves. A computer system is built in such a way that it can only understand 0's and 1's. Start thinking though English language is built with 26 letters it is sometimes hard to understand. Imagine how can a human brain understand language made only with 0's and 1's. So, we the coders started building languages that human find it easy and the machine can understand. Later on, fine tuning the languages and versing them error free, the bridge is made. Software is not only building languages. We guide the computer what to do and not, we act as mediators converting the human given instructions/ requirements to computer's language so that computer can interact with the human. We are not just coders. So we are the converters of this complexity to simplicity.

Whole world want their works to be done easily and most importantly, none tolerates any minor glitches and are most of the time impatient when it is related to work with technology related gadgets. They start complaining over small things. On the other hand everyone say our job is easy and we do nothing but sitting idle all the time. We strive for hours and days to make it easy for you. But still I can't understand why so called coders are always taken easy. We build up a logical world for you and still you call us worthless!! Let's say we deal with a bank related application. The whole world wants all the related transactions and operations to be done within seconds without any glitches and no interruptions just by sitting at home. This is all the freedom that we provide to you just by our so called worthless jobs. We build the interfaces between users and the bank officials so that it becomes effortless for users as well as banks. We take care of everything...Literally everything!! We take care of your crucial information from exposing to the world, we keep your money safe and notify you even if a single paisa move out of your account. Remember its just one of our works, we do a lot more as such.

I remember a quote by Steve Jobs - "*Everyone in this country should learn how to program because it teaches you how to think*". I feel this one quote answers many of today's questions and comments. We don't simply sit and type statements. We do it to achieve a purpose and most importantly we think to draw out some inferences before we actually start our work. Our brains visualize things that do not actually exist and try to bring them to reality through our coding. Seriously, coding brings out the critical and logical thinking capabilities of a person. We build up many systems before we actually produce to the world.

We know computers and smartphones outrun the working of many electronic gadgets today. Everyone say it is good

that all other applications are handy if one smartphone is in our hands but, no one ever understands the struggle and hardwork behind this. We build up all these application softwares and integrate into a single system to make it easy for you without even a single bug and you still pull us back! We strive for your security, convenience, and a lot more. We do all of it for you and you still don't recognize us..What a pity... We learn all the brand new technologies just to make your life easy and to provide you a smooth experience even then you call us worthless...

Not to forget, the transition through electronic world is a collective work. All the domains go hand in hand. Nothing is deprived of appreciation. We are not less than anyone. We do have our own recognition and importance. People from software background has nothing to do if there are no systems to work with. Similarly, building systems won't do if there is proper software defined. So nothing is easier or tougher. Reminding back it's the perspective that defines everything. Probably, the only difference it can make is that we work with logical world building things logically without much physical stress while others work physically. Believe me sometimes even we drain out but mentally. You manage tangible things but we produce intangible world. Despite of all these work we are called easy or sometimes lazy workers. Let me tell you coding isn't that easy as you think. May be joining cables, or parts or mixing concrete at most precision is your work, but at least making the computer manufactured by you wish you a simple "hello" needs the utmost precision, we can't miss a single ';'. We do have things to do.

Finally, to answer the statement Coding is an easy job... Its not our job easy we make your life easy. For you its simple to send a message through a phone, but transmitting that simple message safely involves a thousand lines of struggle. That easy our job is. Though we work logically, we do not compromise your safety, your information security not even at a single point. That dedicated our job is. Again the fun fact is most of those that mock at us for being only coders end up trying for a so called software job at the end.!!LOL!!. Reconsider again no one is deprived of what they do. Whether it is a software job or a core one, we all see success only when united and work together. We, the so called CODERS are always proud of our job, we do it for you..

To know more on coding practices and technologies you can refer,

References:

[1] Li, X. and Prasad, C., 2005, October. Effectively teaching coding standards in programming. *In Proceedings of the 6th conference on Information technology education* (pp. 239-244).

[2] Roth, R.M., 2006. Introduction to coding theory. *IET Communications, 47*(18-19), p.4.

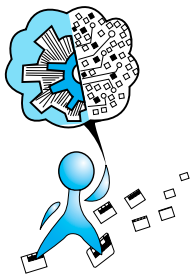
[3] Gasiba, T.E., Lechner, U., Pinto-Albuquerque, M. and Fernandez, D.M., 2020, December. Awareness of Secure Coding Guidelines in the Industry-A first data analysis. In 2020 *IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 345-352). IEEE.

[4] Chi, H., Jones, E.L. and Brown, J., 2013, October. Teaching secure coding practices to STEM students. In *Proceedings of the 2013 on InfoSecCD'13: Information Security Curriculum Development Conference* (pp. 42-48).

[5] Gasiba, T.E. and Lechner, U., 2019, September. Raising Secure Coding Awareness for Software Developers in the Industry. In 2019 *IEEE 27th International Requirements Engineering Conference Workshops (REW)* (pp. 141-143). IEEE.

About the Author

Motupalli Vijaya Sai Priyanka from Bhimavaram, A.P. She completed her Bachelors in Informtion Technology department from JNTUK-UCEV Vizianagaram. She 8.89 CGPA during my college. She got placed in companies like Larsen & Toubro and few others. She selected L&T and currently working in Storage& Backup domain as Graduate Engineer Trainee for Data Center Business, L&T Ltd. She is good at skills like C, C++, Python, DBMS,SQL etc..



Algorithm: The essence of Computing Science

Tirimula Rao Benala

Faculty in the Dept. of Information Technology at JNTUGV, College of Engineering, Vizianagaram.
Email: btirimula.it@jntukucev.ac.in

Algorithms are step-by-step instructions for solving a particular problem. If your algorithm involves significant amounts of input data, complex operations, or both, you need to build clever algorithms that computers can process quickly. The myth is that we do not need algorithms for aspirants to become software developers and full-stack engineers. We can use existing face recognition algorithms, Dijkstra's shortest path algorithm, and external libraries to solve real-time problems. However, MAANG companies like Google use tailor-made shortest path algorithms to find the shortest path between source and destination for Google maps. Similarly, Meta uses sophisticated algorithms for friend recommendations on Facebook. Hence, we must adjust the existing library or build code from scratch. It is not correct to judge a candidate by technology or language ability. Therefore, big companies prefer asking algorithms questions rather than technology or language. Algorithms help test the quality of candidates. Adapt specific solutions to specific situations through new approaches to newly seen problems and confidently present solutions to teams with different perspectives.

In computer science, data (treated as singular, plural, or mass nouns) is any sequence of one or more symbols. Digital data is data represented by a binary number system of 1s and 0s rather than an analog representation. In modern computer systems (since 1960), all data is digital. Data representing quantities, characters, or symbols on which a computer performs operations are stored and recorded in magnetic, optical, electronic, or mechanical recording media and transmitted in the form of digital electrical or optical signals. Digital data is often stored in relational databases, such as spreadsheets and SQL databases, and can generally be represented as abstract key-value pairs. Data structures can store different data types, such as numbers, strings, and other data structures. In computer science, a data structure is a data organization, management, and storage format typically chosen for efficient access to data. More specifically, a data structure is a collection of data values, relationships between them, and functions or operations that can be applied to data. This is the algebraic structure of data (Wikipedia). Algorithm designers must choose appropriate data structures to design efficient algorithms.

Algorithm design is a method or mathematical process for problem-solving and algorithm development. Algorithm design is part of many solution theories in operations research, Dynamic programming, and the divide-and-conquer principle. Techniques for designing and implementing algorithmic designs are also known as algorithmic design patterns, and examples include the template-method pattern and the decorator pattern.

Resource efficiency is one of the most important considerations while designing algorithms (runtime, memory usage). Using the Big O notation, one may show how an algorithm's run-time grows as the size of its input does.

Typical procedures for creating algorithms:

1. issue description
2. creation of a model
3. Information about the algorithm
4. constructing an algorithm
5. examining the algorithm's accuracy
6. a review of the algorithm
7. application of an algorithm
8. program evaluation
9. preparing documentation

Algorithmic effectiveness [4]:

When developing a resolution to an issue, we frequently want to gauge how "quick" the solution is. Unfortunately, many factors influence how quickly your software runs. The hardware of your computer, the dataset you're using, other applications running on it, and even the temperature of your room can all impact how quickly your program runs. Thus, while using timers to assess program speed is helpful, it is not necessarily a valid technique to contrast different implementations.

Big-O notation: Context

Big-O notation is a more theoretical method of comparing different algorithms. Big-O notation is a mathematical notion that provides a rough upper bound on how long an algorithm will take to execute based on the size of the

dataset it uses. Read this article if you're curious about the notation's precise definition. For the time being, the emphasis is on making the notation simple to comprehend and efficient to apply.

Big-O Notation

It is helpful to know what is vital and what is not essential before using Big O notation. There are a few guidelines to follow while using the notation:

1. Constants are not considered
2. the term with the most rapid growth is considered

Considering those guidelines, the following statements are equivalent:

O(90 * n) → O(n)

O (0.231) → O(1)

$O(4*n^2 + 3n + 2) \rightarrow O(n^2)$

O(n*log(n) + 5000 * n) → O(n * log(n))

In each example, constant values are changed to 1, and other smaller components are not considered.

The mathematical definition of Big O is as follows:

Let f(n) and g(n) be positive functions, and let the set of functions be denoted by O(g(n)). f(n) is a member of the set O(g(n)) for which the ratio f(n)/g(n) remains below a constant c when n becomes very large.

Example #1: f(n) = 50n is O(n), as the ratio of f(n) to g(n) = n is always c=50.

Example #2: 4n-3 is also in O(n) because the ratio between f(n) = 4n - 3 f(n) = 3n2 and g(n) = n remains at or below c = 4.Z

Complexity Analysis [6]:

What is the Time Complexity of the summation of n numbers?

Programmer perspectives:

Input: **n**

Output: **1+2+3+...+n**

```
Algorithm #1: function summation(n) {
var sum = 0; ——→ 1 fundamental operation
for (var i = 1; i < n; i++) {——→ n iteration
  sum = sum+i; ——→ 2 fundamental operation * n
 }
 return sum; ——→ 1 fundamental operation
          }
```

Time Complexity: number of fundamental operations

Fundamental operations:

a+b, a-b, a×b, a÷b, a=b, return, bitwise operations, Comparision operations

Algorithm #1 has 2*n*+2 fundamental operations

Time Complexity: O(2n+2) implies O(n)

Algorithm thinking perspective:

Input: **n**

Output: **1+2+3+...+n**

```
Algorithm #2: function summation(n) {
                return n * (n+1)/2;  ←— 4 fundamental op.
              }
                                      doesn't depends on n
```

Four fundamental operations, independent of input size n, are required.

Consequently, the time complexity is: O (1)

Algorithm #1 has a time complexity of O(n), whereas Algorithm #2 has a time complexity of O (1).

Conclusion:

CRUD applications require no algorithmic knowledge. However, in order to implement an innovative concept, we must design complex algorithms by modifying existing algorithms or from scratch. Thus, algorithms play a crucial role in developing innovative, cutting-edge products such as google maps, Amazon's product recommendation system, and Netflix's movie recommendation system.

References

[1] Algorithm - Wikipedia. (2022, July 1). Retrieved September 10, 2022, from https://en.wikipedia.org/wiki/Algorithm

[2] Data (computer science) - Wikipedia. (2022, January 14). Retrieved September 10, 2022, from https://en.wikipedia.org/wiki/Data_(computer_science)

[3] Data structure - Wikipedia. (2018, November 6). Retrieved September 10, 2022, from https://en.wikipedia.org/wiki/Data_structure

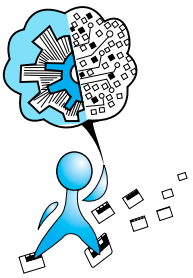
[4] School, R. (n.d.). Free JavaScript Computer Science Course | Rithm School. Retrieved September 10, 2022, from https://www.rithmschool.com/courses/javascript-computer-science-fundamentals/introduction-to-big-o-notation.

[5] Learn Algorithm Fundamentals on Brilliant. (n.d.). Retrieved September 11, 2022, from https://brilliant.org/courses/computer-science-algorithms/the-speed-of-algorithms-2/big-o-formally/3/

[6] Big O Introduction. (n.d.). Retrieved September 11, 2022, from https://rithmschool.github.io/function-timer-demo/

About the Author

Tirimula Rao Benala obtained Ph.D. from JNTU Kakinada. He has been working as a faculty in the Department of Information Technology at JNTUGV, College of Engineering, Vizianagaram, for the last ten years. He has published about 35 refereed journal and conference papers. Currently, he is heading the department. He has total teaching experience of 20 years.



Analysis of Auto Encoders for IOT Botnet Detection

Sai Vasanthi Mandala

Pursuing Masters in the Stream of Data Science at JNTU Gurajada Vizianagaram Email: saivasanthimandala@gmail.com

IoT Botnet detection utilizing deep-learning algorithms has been extensively researched. Network attacks are continually and dramatically changing, exhibiting new patterns. The use of various autoencoders for IoT Botnet identification has recently increased to reliably and quickly identify unknown attack types (such as zero-day assaults) and to lessen the load of the time-consuming labeling operation. It takes a lot of time and effort to identify the ideal model architecture and hyperparameter values of the autoencoders that produce the best detection performance, even though the autoencoders are effective in detecting new sorts of attacks. This may be a barrier to autoencoder-based IoT Botnet detection in real implementations. We thoroughly examine autoencoders using the benchmark dataset N-BaIoT to overcome this issue. Using a basic autoencoder model, we compare various model architectures and latent size combinations. The outcomes show that an autoencoder model's latent size can significantly affect the IDS performance.

The global value of the Internet of Things (IoT) devices is predicted to range between $4 trillion in the low estimate and $11 trillion in the high estimate by 2025. The growth of IoT technology has brought an increasing number of gadgets into our lives, making system security a top priority. Many of the modern gadgets we use daily, including smartphones, wearable technology, health monitoring devices, etc., produce enormous volumes of private data but have little to no built-in protection. Even if the internet is already difficult to secure, the work is made more difficult by the extra unsecured IoT devices. Any internet-connected device, including mainframes for businesses as well as smartwatches and household smart kitchen appliances, can be compromised by botnets. The free accessibility of the source code for IoT botnets like BASHLITE and Mirai has encouraged hackers to experiment with IoT malware. The IoT malware known as Mirai has led to a revival in IoT malware and has been the cause of significant DDoS assaults. The Mirai botnet, as well as its variations and copycats, essentially served as a wake-up call for the industry to strengthen IoT device security.

The goal of botnets is normally to infect as many devices as they can, and complex botnets can even self-replicate and adapt their behavior to identify and infect devices on their own. Consequently, botnets are incredibly challenging to find. The fact that botnets hide on devices with minimal impact on the device's functionality is another reason why they are challenging to find and contain. For instance, neither the typical user nor a small business may be aware that a security camera is a component of an active botnet. Therefore, it is crucial to distinguish botnets from IoT device traffic. In this, we classify botnet traffic in the IoT context using the dataset from. Nine commercial IoT devices that were attacked by the Mirai and BASHLITE botnets provided the data for this dataset, which contains genuine network traffic information. Three classifiers, Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF), Autoencoder, are used to assess the data, which is then categorized by a botnet, attack, and device.

The security of network systems and information assets from network attacks is crucial, and there are numerous techniques for accomplishing this. Of these, source authentication utilizing cryptography with public keys and message authentication can be used to secure network systems. Now it is simple to determine whether network traffic originates from a reliable source or not due to these encryption techniques. As just a result, we are capable of filtering out malicious traffic from shady sources. Post-quantum public key cryptography is being developed to take its place because conventional public key cryptography will have security issues with the introduction of quantum computing. Intrusion detection systems are a common and efficient defense against network attacks to overcome this constraint.

IoT Botnet detection has recently been established by several research teams employing autoencoders, a generative deep-learning model made up of an encoder and a decoder. When M > N, the encoder transforms an input M-dimensional vector into a latent vector represented as an N-dimensional vector, which

the decoder then reconstructs back to the original M-dimensional vector (see Figure 1). Any input should be recovered as close as feasible to the taught usual patterns by an autoencoder-based IoT botnet detection system that was trained with only typical traffic data. Therefore, if an input instance's reconstruction error exceeds a certain threshold, we can label it as an assault; If not, the input instance can be categorized as normal. In this way, an autoencoder-based IoT Botnet detection system can identify unidentified attack types when their patterns diverge from the recognized typical patterns.

To the best of our knowledge, we are the first to use autoencoders to analyze IoT network traffic for anomalies as a comprehensive way to identify botnet attacks. Even in the broader field of network traffic analysis, autoencoders have not been applied as fully automated standalone malware detectors, but rather as preparatory tools for feature learning or dimensionality reduction, or at most as semimanual outlier detectors that heavily rely on human labeling for subsequent classification or further inspection by security analysts.

We conduct an empirical evaluation with real traffic data, gathered from nine commercial IoT devices infected by real botnets from two families, in contrast to previous experimental studies on the detection of IoT botnets or IoT traffic anomalies, which relied on emulated or simulated data. We look at two of the most prevalent IoT-based botnets, Mirai and BASHLITE, which have already proven their destructive potential.

Previous IoT-related detection studies generally concentrated on the early stages of propagation and communication with the C&C server when examining the operational phases of botnets. However, given that botnet attacks continue to mutate every day and become more complex we predict that some of these mutations may eventually succeed in getting around current early detection strategies. Moreover, when connected to external networks, mobile IoT devices could become infected. For instance, when their owners arrive at airports, smartwatches may connect to suspicious free Wi-Fi networks. As a result, simply monitoring organizational networks to spot early infection signs is insufficient. As a result, we concentrate on the phase of a botnet operation when IoT bots start initiating cyberattacks. In that regard, our solution offers a final degree of security protection. It quickly recognizes IoT-based attacks and lessens their damage by sending out an instantaneous notice that suggests isolating any affected devices from the network until they are cleaned up.

A key distinction between host-based and network-based approaches is noted among the suggested botnet detection methods. Because we cannot rely on IoT manufacturers to install designated host-based anomaly detectors on their products, there is limited access to some IoT devices (such as wearables), so the installation of software on end devices cannot be enforced, and the limited computation and power of most IoT devices place restrictions on the complexity and efficiency of host-based techniques, we believe host-based techniques are less realistic for detecting compromised IoT devices, In the enterprise scenario we assume, where various and numerous IoT devices connect to the organizational network, a single non-distributed solution is preferred. This is because multiple distributed solutions could potentially consume energy and computation from the devices and harm their functionality.

The IoT area is not the only one for which a hierarchical taxonomy of network-based botnet detection methods is presented. One of the sources of detection examined in this study is honeypots. For gathering, comprehending, describing, and tracking botnets, honeypots are frequently utilized. They do not always help with identifying compromised endpoints or the attacks coming from them, though. Additionally, honeypots typically need a sizable expenditure in the acquisition or replication of real devices, data inspection, signature extraction, and mutation monitoring. Accordingly, regular networks serve as a secondary source of detection, and network intrusion detection systems (NIDS) use pattern matching to identify indications of malicious activity by continually and automatically monitoring traffic data. These patterns may rely on honeypot signatures, DNS traffic involving a putative C&C server, data mining of traffic anomalies, or hybrid strategies. Because linked appliances are often task-oriented, we found that the anomaly-based method is most suited for identifying corrupted IoT devices (e.g., specifically designed to detect motion or measure humidity). As a result, they run fewer and maybe simpler network protocols and have less variable traffic than PCs. As a result, spotting changes from their typical patterns ought to be more precise and reliable.

Numerous detection techniques were reviewed; however, no autoencoders nor artificial neural networks were cited. However, they differ from our strategy, have nothing to do with the Internet of Things, and frequently have no direct connection to botnets. Such publications within the broader subject of cybersecurity have been published more recently. As an illustration, and used shallow autoencoders for initial feature learning and dimensionality reduction, Random Forest, Deep Belief

Networks, and Softmax for classification, and then for final fine-tuning. Although outlier detection was included in autoencoders, security analysts still needed to explicitly classify data for later supervised learning. The authors use deep learning to analyze system logs to find insider threats, which is more similar to our method. Unlike us, they rely on further manual scrutiny and use DNNs and RNNs (LSTMs).

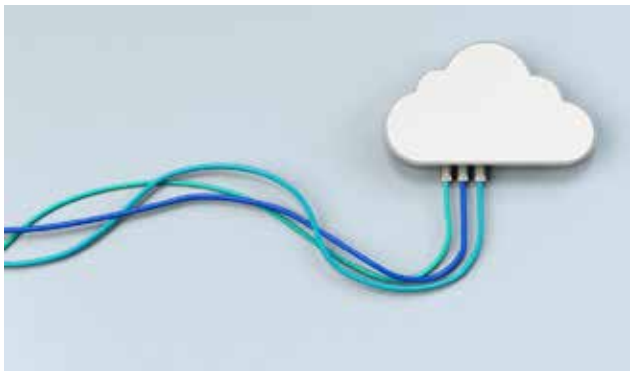
Although the majority of the IoT devices in a test set gave the autoencoders in our trials an FPR of zero, the variation in FPR among the remaining IoT devices prompted us to further study our data. The Philips B120N/10 baby monitor had the highest FPR in comparison to the other devices, and it also generated the most traffic (see Table 3), thus one may anticipate that the huge number of training examples would lead to more reliable machine learning models. However, this device also has the widest range of capabilities due to the presence of many sensors for ambient light, temperature, and humidity, a two-way intercom feature, motion detection, and audio detection. This may make it more challenging to observe it acting normally, which could lead to more classification mistakes in subsequent observations.

In light of this, we postulate that different IoT devices have varying degrees of difficulty in capturing the usual traffic behavior and that this challenge could be due to (1) the capabilities of the device and (2) network communications typically, results. then, the simplicity of IoT device baseline behavior establishment facilitates Attack detection by anomaly detection. As such Finally, intriguing queries are raised:

- Can the traffic behavior of IoT devices be predicted? quantified?
- Can there be a correlation between the degree of predictability and the features that might be static or dynamic, such as the number and type of sensors, memory size, and operating system? (For instance, the volume of distinct destination IPs per hour, variation of the ratio of incoming to departing traffic) be formalized?
- Can the impact of these characteristics be ranked? this degree of predictability?

We assume that performance metrics for anomaly detection can be easily translated from the predictability of traffic behavior. For instance, an IoT device with high traffic predictability would highlight any aberrant action, causing the TPR to rise and detection times to shorten in this scenario. From the (harmless) training set, we retrieved static and dynamic features for empirical validation. Afterward, we developed regression models to examine the impact of these features on the test set's average FPR and detection times for the four detection methods we considered. Figures 2c and 2d show our initial conclusions using the characteristics we felt were most important. Figure 2c illustrates how an increase in inbound traffic variability results in a bigger average FPR (p-value=0.019). This makes sense because less predictable situations frequently show up as unusual (but benign) traffic behaviors that are mistakenly labeled as abnormal. Figure 2d demonstrates how longer detection times are encouraged (pvalue=0.001) by an increase in the maximum amount of inbound traffic. Lower predictability causes larger ws* (more instances for majority voting), which in turn causes longer detection times as we optimize ws* to achieve 0% FPR on DS_{opt}.

The Review



Corporate newsletter

– By Shir Rosenstein

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They can be printed or emailed and are an excellent way to maintain regular contact with your subscribers and drive traffic to your site. Type the content of your newsletter here.

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog.

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They can be printed or emailed and are an excellent way to maintain regular contact with your subscribers and drive traffic to your site. Type the content of your newsletter here.

Work with the industry's best

– By Taylor Phillips

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They can be printed or emailed and are an excellent way to maintain regular contact with your subscribers and drive traffic to your site. Type the content of your newsletter here.

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. Type the content of your newsletter here.

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They can be printed or emailed and are an excellent way to maintain regular contact with your subscribers and drive traffic to your site. Type the content of your newsletter here.

The observer

– By Chanchal Sharma

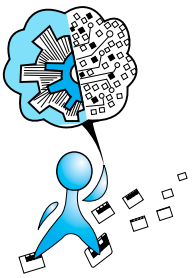
Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They can be printed or emailed and are an excellent way to maintain regular contact with your subscribers and drive traffic to your site. Type the content of your newsletter here.

Newsletters are periodicals used to advertise or update your subscribers with information about your product or blog. They are an excellent way to maintain regular contact with your subscribers.

About the Author

Sai Vasanthi Mandala worked on power domain projects. Process in KYRGYZSTAN is to test the modules like New Service Connection, Customer Support, Metering, Billing, Payments, Meter Management, Configurations and also the same done in another project 'OSHEE' (Albania) in APDCL (Assam Power Distribution Company Limited) is to test the modules like MDMS (Meter Data Management System), FEP, OSM (Outage Management System), DSM (Demand Side Management), PQM(Power Quality Management), NMS, CP(Customer Portal).

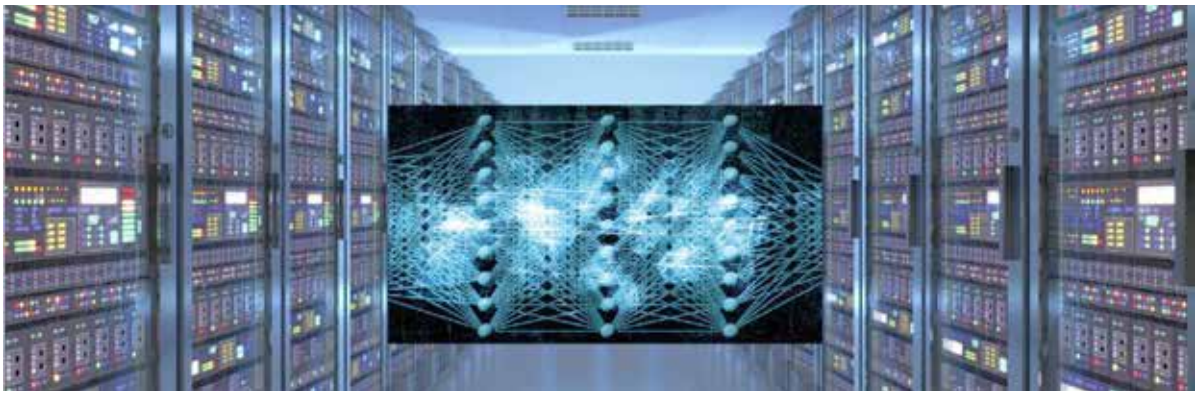
She worked as freelancer in online platforms like User Testing, U-test. Also have good knowledge and Experience in Digital Advertising, Digital marketing. She worked as a Subject Matter Expert in the stream Computer Science.



Predicting next item interaction with session aware and session based Recommendation Systems

Shaik Mohammad Yunus Ansari

Pursuing Masters in the Stream of Data Science at JNTU Gurajada Vizianagaram . Email: yunus00939@gmail.com



Picture Courtesy: https://telanganatoday.com/hyderabad-ideal-to-set-up-data-centres
https://voith.com/corp-en/news-room/stories/how-artificial-intelligence-works.html

Recommender System

We live in a digital world where we can access to varieties of information and every application and machines are being automated with AI computer algorithms. These algorithms are used in various online applications like Ecommerce, Music, Video-Streaming etc, the main advantage of using Artificial Intelligence in these fields helps businesses to gain profits by understanding the customer or user behaviour, understanding latest trends, incorporating best strategies into their applications to grow their business.

Recommender systems plays an important role in an online e-commerce websites and media streaming websites which are widely used where it helps to find the individual user interests and providing suggestions. Every recommendation system main goal is to provide personalized recommendation which is relevant to the user interest.

These recommendations can be attained by different types of methods where every system works with an input to provide a definite output. Every business uses their own strategy which personalizes the recommendation system to understand the individual user-interest by different factors like age, gender, demographic location, past user interaction data etc, because there are more than 7 million of people possessing different behaviour and it is little tricky to predict the exact behaviour but with these input factors we can assume that this type of users may have some similarities. For example, a new user has signed up to an online shopping application by not providing complete information. Being a new user there will be no past information which makes system to experience cold start problem which means not understanding which suggestion to be provided.

With the few interactions with the items the system recognizes the pattern and can provide suggestions with the help of Collaborative-filtering technique.

Ex. Being anonymous. By providing incomplete information to the recommender, it cannot be able to provide accurate suggestions and there are also several other factors which recommender system is affected like concept drift, data sparsity etc. Even though, the user doesn't provide any input data, or effected by various factors the successful recommendation system has to provide suggestions precisely to user interest.

Deep Learning in Recommender Systems

Recently, deep learning is applied to recommender systems because these recommender systems are applied almost all the online websites which is having different types of data format. There are two types of data ie., structured and unstructured data mostly of the information available online. Structured data is textual data like reviews, tweets, gradings, ratings etc and unstructured data is image data

like social posts, product images etc, it is impossible to draw suggestions for these types of information so deep learning models plays an important role for those types of recommender systems. Deep learning is typically regarded as a subfield of machine learning.

Content Based Filtering

Based on the previous actions of the user or feedback given to the product/item, Content based filtering uses item features with which similar items or products are recommended.

Collaborative Filtering

This filtering finds similarities between users and items ie., user to user, item to item, item to user simultaneously to recommend the products to the user.

Hybrid Filtering

It is the combination of both content and collaborative filtering

Types of similarity measures: Using distance metric the similarities are identified and those are categorized relevant or irrelevant using distance. The distance can be calculated by using Minkowski distance, Manhattan distance, Euclidean distance, Cosine similarity, Pearson coefficient, Jaccard Similarity, Hamming distance.

Deep learning is commonly defined as learning deep representations or learning several levels of representations and abstractions from data. It improves a differentiable objective function using a stochastic gradient descent variation (SGD). Both supervised and unsupervised learning tasks have seen significant success using neural architectures. Some of the deep learning techniques are Multilayer perceptron, CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), Auto Encoders, DRL (Deep Reinforcement Learning) etc.

Sequential recommendation system:

These systems rely on traditional user-item rating matrix as an input. The goal is to predict immediate next user-item or point of interest that user will interact from the time stamped rating matrix of sequence of events in data.

Session-based recommendation system:

This system predicts the next user action present on the ongoing interactions of current session. In this system users are anonymous which deals with first time users or users who are not logged in. The input of these systems are recorded user interactions arranged according time ordered.

Session-aware recommendation system:

This system is also known as personalized session based because the main goal is also identical but users are known in this system, ex: users who are logged in. In this system recommendations are personalized with user past behaviour which can be beneficial for accuracy.

Several researches were conducted to understand what represents the state-of-the-art by evaluating different neural approaches. Comparisons were made by comparing newly proposed session-aware with existing session-based models which do not use a consistent set of baseline algorithms, so still research is scattered. The main idea is to combine the user's long term and short term intents which can relate to the better accurate suggestion so we have compared five recent neural models like HGRU4REC,IIRNN(Inter-session and Intra-session Recurrent Neural Network),SHAN(Sequential Recommender System based on Hierarchical Attention Network),NCSF, NSAR(Neural Session-Aware Recommendation)with set of existing neural and non-neural approaches to session based and heuristic extensions like Extend, Boost and Remind of session based techniques which uses interactions in ongoing sessions. Nearest neighbors in session based are also used because they are very competitive.

Four datasets Cosmetics, Lastfm, Retail, Xing which are publicly available were considered in this experiment. Surprisingly, heuristic extensions of existing session-based algorithms were best performing. In many cases nearest neighbors outperform recent session aware models although they do not use long term users intents.

With the analysis of several neural, non-neural and nearest neighbors baseline results, the session aware which combines users' long term and short-term intents can be able to provide accurate suggestions if they were developed and still has huge potential to be state-of-the-art in predicting next user item interaction.

References

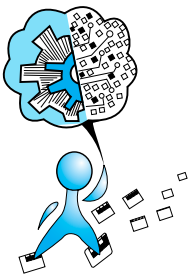
[1] https://www.sciencedirect.com/science/article/pii/S0020025521005089

[2] https://analyticsindiamag.com/collaborative-filtering-vs-content-based-filtering-for-recommender-systems/

[3] https://arxiv.org/pdf/1707.07435

About the Author

S M Yunus Ansari is currently pursuing his masters in the stream of Data Science at JNTU Gurajada Vizianagaram and completed his bachelors at KL University, Guntur. His fascination is towards practices in neural structures of deep learning in the domain of recommender systems and his current research interest lies in Data Science, Machine Learning and Deep Learning.



Hybrid Similarity Metrics for Neighbor Based Collaborative Filtering Performance Improvement

Mouneswari Pentakota

Pursuing Master of Technology in the stream of Data Science at Jawaharlal Nehru and Technology University, Guarajada, Vizianagram. Email: mounieswari123@gmail.com

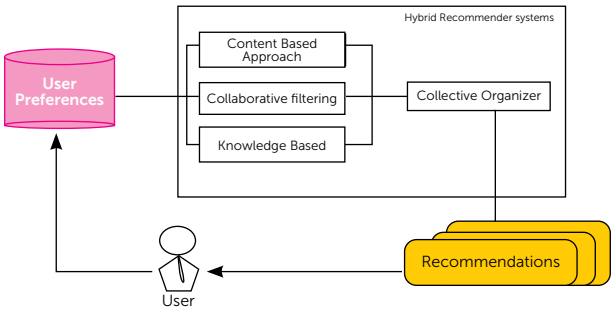
Keywords: Collaborative filtering, K-nearest-neighbor, Recommendation system, Item-based, Memory-based, Similarity measurement

Recommendation systems are a type of information filtering system that uses historical data to forecast a user's opinion or preference on a topic or item, allowing them to receive customised recommendations. They're common on e-commerce sites like amazon.com, online movie streaming services like Netflix, and social media sites like Facebook. With such a great number and variety of products, a recommendation system could also assist streaming services or online sellers in providing customers with recommendations according to their preferences.

This could improve the user experience when looking for products or services, leading to increased purchases, movie viewing, or service subscriptions. For example, statistics collected over three weeks in the summer of 2001 revealed that between 20% and 40% of Amazon purchases are attributable to recommended things that are not among the shop's 100,000 most popular items, while 60% of Netflix movies are chosen based on personalised recommendations. Furthermore, a recommendation system could produce additional money by introducing shoppers to new categories, in addition to greater direct revenue. As a result, a recommendation system might have a big impact on a company's revenue.

Note that while a 1% improvement in average MAE(Mean Absolute Error) and RMSE(Root Mean Squared Error) may seem insignificant, it can make a big difference in a user's ranking of the "top10" most suggested movies. As previously stated, recommendation systems can take one of three approaches: content-based filtering, collaborative filtering, or a combination of the two.

Hybrid Similarity metrics is implied to overcome the constraints of their individual approaches and increase performance such as prediction accuracy and scalability, this strategy combines two or more techniques or uses other techniques such as deep learning or clustering. However,it adds to the computational complexity, requiring more time and resources. Deep neural networks have had a lot of success in computer vision and natural language processing in recent years. Graph Convolutional Networks (GCNs) have also been successfully applied to recommendation systems.



Proposed Approach is to quantify the similarity between items, we suggest a new similarity measurement. Unlike standard CFs, which rely solely on rating-based similarity measures like adjusted cosine, Pearson correlation coefficient, or structural similarity measurement, we propose combining the two. Measures of resemblance based on ratings. The MovieLens and Netflix datasets are selected for comparison because they are the most extensively referenced in the literature. We evaluate and contrast two benchmarks for evaluating prediction accuracy: MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error), and present experimental results on three widely used datasets: Netflix, MovieLens 100K, and MovieLens1M

On the MovieLens 100 K, 1 M, and Netflix datasets, our technique beats state-of-the-art collaborative filterings in terms of lower MAE with 1/3 to 1/2 the number of neighbours as compared to standard memory-based CFs. On the MovieLens 1 M dataset, memory-based CF with the suggested similarity measurement uses 1/2 to 1/39 wall time compared to state-ofthe-art model-based CFsFor non-cold start users, our method yields a 3% lower MAE and RMSE than traditional memory-based CFs on the MovieLens 100 K dataset.

Characteristics and Challenges are We concentrate on memory-based collaborative filtering since it may be

used to any relational data without understanding what it contains. However, collaborative filtering faces several fundamental obstacles in predicting an accurate rating in real time. Sparsity Cold Start Problems are one of the most common problems that data sparsity can cause. Because collaborative filtering systems produce suggestions based on previous preferences, new users must rate a significant number of items for the recommendation system to learn their preferences and make valid recommendations. When users have just evaluated a few things, collaborative filtering systems are often unable to produce accurate recommendations.

Scalability Traditional collaborative filtering approaches will have scalability issues as the number of people and objects grows very large. Model-based techniques would find it difficult to react in real time to fresh user ratings in order to generate an updated suggestion because they train on the complete dataset. however, most users have only reviewed a small percentage of the total number of products, and memory-based techniques can react to new ratings and make predictions in real time, even for massive datasets. While using dimensionality reduction techniques like SVD to minimise scalability issues, model-based approaches suffer from computationally expensive matrix factorization and may lose essential information in the process

Curse of Dimensionality In order to find related objects or users, collaborative filtering must calculate similarities between them. Because there are many users and things, the pairwise similarities are estimated in high dimensions. The Hughes Phenomenon states that as dimensionality rises, the predictive power of a certain size of training samples decreases (Hughes, 1968).When all users or objects have the same similarities, memory-based CF is unable to locate the most comparable items or users, and so cannot provide a trustworthy forecast. The quality of hubness In the nearest neighbour lists of other items, certain items appear more frequently than others. Those things are typically popular high-rated items that do not contribute any personal preference information to suggestions because they may be enjoyed by a large number of users. They can act as noise, preventing memory-based CF from generating precise predictions.

Improvements over using rating or structural similarity alone

Compensate unpopular but similar items: We do not penalise similarities between items because of their unpopularity if both items have been evaluated by the majority of users. Instead of utilising a set global shrinkage factor to punish pairs of items with a small number of co-rated users, the structural similarity measurement compensates unpopular but heavily co-rated items by applying a local ratio of |Cij| to the local possible co-rated users |ii| and |ij|. We estimate the prediction accuracy to be higher than if we only used rating-based similarity measurement

Reduce Hubness Big hubs are those generally liked popular goods when calculated based on rating-based similarity measurement alone. Those goods don't add much to personal preferences, yet they're frequently mentioned as neighbours. The hubness of employing local ratio structural-based similarity weighted rating-based similarity should be lower than rating-based similarity alone. When just structural data is used to calculate similarities, huge hubs form when one user reviews unpopular goods that are rarely rated by other users. As a result, structural similarity will be high.

Full Picture We can anticipate whether a user likes or dislikes an item based on his or her opinions on highly connected neighbour items by merging the two. The strongly connected items are found using structural similarity assessment, and the user loves the item is determined using rating-based similarity measurement.

References :

[1] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17, 734–749. Aggarwal, C. C. et al. (2016). Recommender systems. Springer. Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. Information Sciences, 178, 37–51.

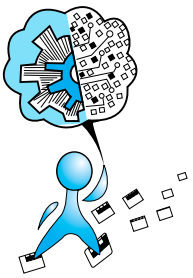
[2] Bennett, J., Lanning, S. & et al. (2007). The netflix prize. In Proceedings of KDD cup and workshop (Vol. 2007, p. 35). New York, NY, USA. Billsus, D., & Pazzani, M.J. (1998). Learning collaborative information filters. In Icml (Vol. 98, pp. 46–54).

[3] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993–1022.

[4] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. Knowledge-Based Systems, 26, 225–238.

About the Author

Mouneswari Pentakota is currently pursuing her masters of Technology in the stream of Data Science and Jawaharlal Nehru and Technological University Guarajada, Vizianagaram. Her fascination is towards problem solving, innovate and research in the field of Data Science. Her works are related as Predictive models to extracts the data the business needs and help analyze the data.



Automated wheelchair using IOT

Eswar Patnala

Faculty in the Dept. of Information Technology at JNTUGV, College of Engineering, Vizianagaram.
Email: eswar.patnala@gmail.com

The Internet of Things (IoT) describes the network of physical objects—"things"—that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet.

IOT has many applications in the real world and the automated wheelchair is one of the example for its applications.

This automated wheelchair enables users such as physically challenged people to move from one place another place. It helps the users to move of their own without any other persons involvement.

For example let us consider some localized areas such as malls, hospitals, bus stop, railway station etc. In these areas the physically challenged people can simply sit and travel from one place to another place by giving a simple instruction.

The instruction can be given through the application and the application is developed in such a way that the instruction can be given both in text and audio format so that both blind and physically disabled person can use the service. Currently some automated wheelchairs are available in the market but we have to operate the wheel chair to travel from one destination to another destination but the automated wheel chair that we have designed will drive automatically once we fixed the destination like a tesla car.

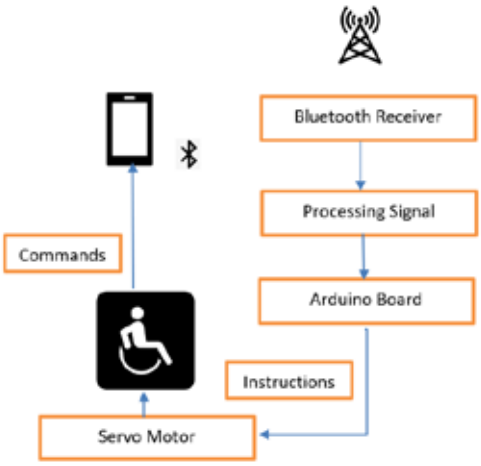


Fig. 1: Process Flow

Work Flow:

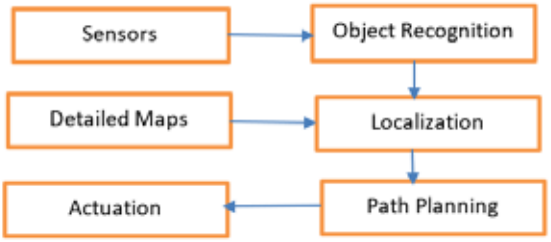


Fig. 2: Work Flow

Sensors:

A sensor is a device that detects input of any kind from the physical world and reacts to it. Light, heat, motion, moisture, pressure, and a variety of other environmental phenomena can all be inputs. The output is typically a signal that is translated into a display that can be read by humans at the sensor location or that is electronically sent over a network to be read or put through more processing.

Object Recognition:

Object detection determines whether or not an object is visible in a picture. It establishes its position within the frame. Although the algorithm utilised to do this task is rather straightforward, it yields effective results. Sensors collect data according to the IoT standard. Instead of sending data across the network or to the cloud, it is more effective to process this continuous stream of data (such as from a video camera) on the sensors.

Cameras carry out object detection along with related activities like object classification and counting. The device's efficiency is increased by performing this preprocessing work locally because extra network traffic is prevented.

Localization:

One of the most alluring IoT applications is localization-based services. They are actually able to gather and send data in order to pinpoint the location of the target because of the deployment of sensor networks. The literature has numerous localization system proposals.

Path Planning:

Any autonomous mobile robot must perform path planning in order to determine the best free path from a given starting place to a specified objective. In this work, a novel approach based on the internet of things (IoT) has been developed; robot platforms are employed for path following, while microprocessor-based gadgets are used for path development. In order to construct a hybrid technique using fixed cell decomposition for the environment and create a map of this environment, two separate algorithms, Bug0 and Potential Field, are combined.

Actuation:

A physical object ("thing") plus a controller ("brain"), sensors, actuators, and networks make up an Internet of Things device (Internet). A machine component or system known as an actuator moves or controls a mechanism or a system. The device's sensors gather information about its surroundings, and control signals are then created for the actuators in accordance with the activities that must be taken.

An actuator is something like a servo motor. They can move to a defined angular or linear location and can be either linear or rotatory actuators. To meet our needs, we can employ servo motors in IoT applications to rotate the motor by 90 degrees, 180 degrees, etc.

Detailed maps:

The process of mapping involves obtaining geographic data such that the user may interpret latitude, longitude, depth, and other dimensions to represent the positions of objects and, optionally, their qualities (such as roads or towers). There are a wide range of potential uses for maps, including in the fields of urban planning and navigation. The application of mapping technology is possible in a variety of industries, including media and entertainment, construction, the healthcare sector, and others.

References:

[1] https://www.jetir.org/papers/JETIR1903F09.pdf

[2] https://www.techtarget.com/whatis/definition/sensor

[3] https://www.researchgate.net/publication/327227837_A_Survey_of_Localization_Systems_in_Internet_of_Things#:~:text=Localization%2Dbased%20services%20are%20among,are%20proposed%20in%20the%20literature.

[4] https://alwaysai.co/blog/using-object-detection

[5] https://ieeexplore.ieee.org/document/8652050

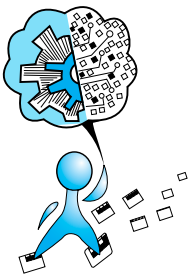
[6] https://www.geeksforgeeks.org/actuators-in-iot/

[7] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3329219

[8] https://www.techtarget.com/iotagenda/definition/Internet-of-Things-IoT

About the Author

Eswar Patnala pursuing PhD from Gitam Institute of Technology, GITAM. He has been working as Assistant Professor in the department of Information Technology at JNTUGV, College of Engineering, Vizianagaram from last nine years. He has published Various International Journals and Book Chapter.



Time Series Analysis of Satellite Images using Pixel-Set Encoders and Temporal Self-Attention

P. AnanyaSahiki

pursuing his Masters of Technology in the stream of Datascience at Jawaharlal Nehru and Technological University Gurajada, Vizianagaram. Email: ananyapechetti@gmail.com

Satellite image time series, Large-scale control of agricultural parcels is a problem that is extremely important from both a political and economic standpoint. In this regard, automated classification of satellite image time series using hybrid convolutional-recurrent neural architectures has shown promising results. We suggest an alternate method that takes advantage of the convolutional layers being replaced with encoders that work with arbitrary sets of pixels to take advantage of the normally low resolution of openly accessible satellite photos.It is a collection of time- and space-varying satellite photos of the same location. A SITS utilises various satellite sources to produce a larger data set with close space between images. The resolution and registration restrictions must be respected in this scenario.

It is possible to recognize how the Earth is changing, to identify the factors causing these changes, and to forecast future changes by using satellite measurements. The combination of data from ecosystem models and remotely sensed data presents a chance to forecast and comprehend the behaviour of the Ecosphere. It is easier to observe accurate spatio-temporal structures in range images with sensors that have high temporal and spatial resolutions. Spectral and spatial dimensions along with temporal elements enable the detection of complex.

For automated crop categorization, practitioners primarily use classic techniques like Randomized Rainforest (RF) and Svms (SVM), which work on handcrafted features. Convolutional neural networks and recurrent neural networks, two deep learning techniques recently used for acquiring spatial and temporal features have significantly improved classification performance. For the purpose of classifying crops, hybrid neural architectures that combine convolutions and repeating units into a single architecture are the most advanced at this moment.

The distribution of spectra so over parcels' full spatial extent can be easily separated using set-based encoders.

About the Author

Ananya Sahiki Pechetti is currently pursuing his Masters of Technology in the stream of Datascience at Jawaharlal Nehru and Technological University Gurajada, Vizianagaram.His enthusiasm towards Datascience and Machine learning engineering motivated to analyse and executive solutions in the field of Datascience.With this background, he is adept at picking up new skills quickly to deliver robust solutions to the most demanding of businesses.

Dynamis Behavior based Customer churn Prediction

Venkata Rajat Kumar Vuppala

Pursuing his Masters in the stream of Data Science at JNTU Gurajada Vizianagaram. Email: vvenkatarajat@gmail.com

Customer behavior plays a vital role in customer service sector to accurately predict the customer churn. Customer churn arises when the customer does not want to continue their relations with the company during certain time frame. As there is increase in competition among customer service sector, it has dominant to predict the Dynamic behavior of the customer. We have adapted advent of the deep learning paradigms namely LSTM and variants of LSTM based models to predict the churn on daily dynamic behavior of the customer. The predictive performance of these models evaluated on dataset collected from mobile operator. The results showed that the daily models significantly outperform previously developed models in terms of predicting churners earlier and more accurately.

Keywords : Churn Prediction; Mobile Telecom; Deep Learning; Dynamic Behavior.

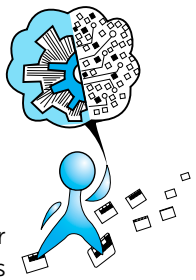
Customer churn refers to the situation where by a customer leaves a service provider. In simple terms, churn is defined as cancellation of subscription by a client to a service they have been using. Churn Prediction is essential in predicting the clients most likely to unsubscribe based on their usage of services like company tariffs, poor customer care or frequent technical issues, helping organizations to focus more on the customers with high risk of leaving.

Customer churn happens when customers decide to not continue purchasing products/services from an organization and end their association. It is an integral parameter for the organization since acquiring a new customer could cost almost 7 times more than retaining an existing customer. Customer churn can prove to be a roadblock for an exponentially growing organization and a retention strategy should be decided in order to avoid an increase in customer churn rates.

The ability to be able to predict that a certain customer is at a very high risk of churning, while there is still some time to do something significant about it, itself represents a great additional potential revenue source for any business. Since the customer is the major source of profit, a method to promptly manage customer churn gains vital significance for the survival and development of any telecommunication company. For many telecoms companies, figuring out how to deal with churn is turning out to be the key for continued existence of their organizations.

In order to learn more about this issue and come up with a workable solution, research and implementations done in the past by other authors were examined. Churn studies are primarily focused on industries with contractual settings, such as telecommunications, banking, insurance, etc., where a consumer must sign a contract in order to use the service. As a result, the company labels a customer who cancels their contract as a churner, as opposed to a customer who keeps using the company's services and is recognized as a non-churner.

The dataset for this research is based on the telecom statistics gathered from the Francisco gallery of bigml. com will be the dataset we use for the majority of our discussion in this paper. This consists of 3333 examples and 20 qualities, along with a final categorization of either "Churned" or "Not Churned" for each client, is based on telecom facts acquired from bigml.com's Francisco



gallery. There are 483 churns and 2,850 non churn customers in the dataset. This dataset is also used to predict the customer's behavior. It gives basic information about the customers service usage as well as information on their membership, both of which are useful for training the base model. Each record is described by the following attributes i.e. churn as class label, area, service calls, evening calls, evening charge, minutes spend in the evening, day calls, day charge, minutes spend in day time, international calls, international charge, minutes spend in international calls and finally it includes night calls, night charge, minutes spend in nights. The dataset is divided into training set and test set. The deep neural network is trained on the training data and tested on the test data.

Previous works on churn prediction focused on predicting churn monthly based on static or monthly dynamic behavior. Two main drawbacks can be drawn from these works;

1. Predicting churn monthly is late for customers who decided to leave at the beginning of the last month because the customer will be identified as churners by the monthly model at the next month.
2. Considering the monthly behavior ignores the changes in customer's behavior over days of the month and this may reduce the discriminative ability of the prediction model, and thus, its predictive performance.

In this study, we attempt to forecast daily levels of customer churn using dynamic shifts in daily activity. As a solution, we used a multivariate time series to represent the customer's daily behavior and addressed the problem of daily churn prediction using this representation.

Meaningful features from timeseries data are extracted and then fed to a traditional machine learning model, such as Random Forest, to make predictions. Deep Learning models like LSTM and its Variants, which include classic LSTM, stacked LSTM, bidirectional LSTM, and GRU, will learn the representative features from the multivariate time series and simultaneously predict churn.

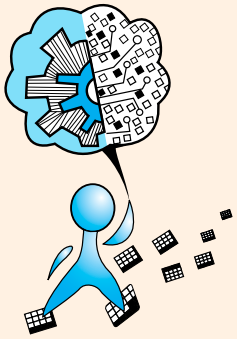
The practical implementation of this study is suggested by our findings, which show that the deep learning models we suggest are operationally effective and forecast churners earlier and more precisely than the machine learning models and monthly models.

About the Author

Venkata Rajat Kumar Vuppala is currently pursuing his Masters in the stream of Data Science at JNTU Gurajada Vizianagaram and completed his bachelors at Raghu Institute of Technology, Visakhapatnam. His fascination is towards latest advancements in the domain of Deep Learning and Machine Learning and his current research interest lie in Data Science, Machine Learning and Deep Learning.

A human carries a Bytes - supported AI Balloon

Keerthana Achalla is currently working as Software Engineer at Bank of America, chennai. She is graduated in the stream of Computer Science specialization in Data Analytics from VIT-AP University. Her interest on art and painting lead to practice of both technical and design fields.



Campaigns using personalization and recommendations to boost revenue

Vikram Gupta

Front-end architect at Adobe Systems. Email: kvikram1988@gmail.com

E-commerce is growing exponentially with companies going ahead with digital-first mindset and with more and more players entering into this segment, the competition is immense. However, businesses that understand personalization are the ones growing at an accelerated rate by creating engaging human connections. Creating unique experiences for large audiences isn't easy. Implementing personalization at scale is complex, involving transformation of your data, content, teams, and technology. But the effort is necessary, given 80% of B2C customers and 84% of B2B customers feel that personalization saves them time by making it easier to find information and settle decisions.

The target is to enable marketers and sales to deliver and optimize highly personalized experiences on any website through recommendations and personalization. To serve the most appropriate content and offers to audiences based on contextual data such as user browsing, search, and product purchase history. To provide an omni-channel solution that improves the visitor experience on any surface or screen customers engage with, including websites, native mobile apps, set-top boxes, kiosks, and more.

An example of such offering is Adobe Target shown below:

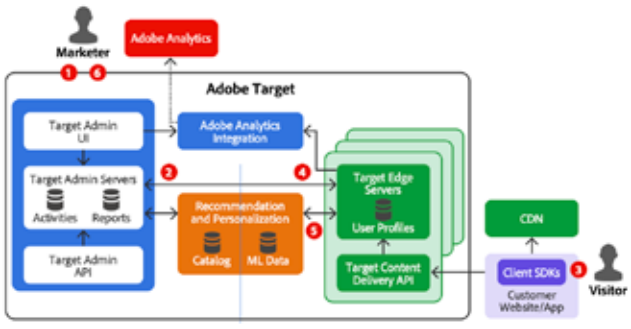


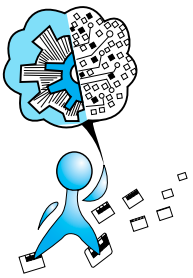
Fig. 2: Adobe Target Data Flow Diagram

Personalized Recommendations uses a combination of sophisticated algorithms to develop recommendations for individual users. Among these are Item-Item Collaborative Filtering, which uses a novel similarity metric that corrects for popularity bias and allows for the exploration of long-tail items in recommendations. We also use content-based similarity algorithms that produce recommendations purely based on the descriptions of items in your product catalogues, with precise controls allowing you to tune the relative importance of different words to produce exactly the recommendations you want. Finally, we use popularity-based algorithms that recommend items based on top viewed, top purchased, and top viewed items among users who share a profile attribute, and others.

In addition to above, the apps has to run A/B tests that optimise the overall website experience for all customers. It helps business owners understand their customers and chose the best workflow for them.

The journey doesn't end on the website, it continues even after the website visit through SMS, emails and in-app notification campaigns.

Campaign is a cross-channel marketing campaign management solution that enables organizations to bring customer data from different systems, devices, and channels into a single profile and deliver timely and relevant campaigns to these customers. With this, companies can understand and define unique customer segments and then craft multi-step, cross-channel



campaigns that make meaningful connections with each customer.

All of the above has to deal with giga bytes of user's data. Profiles built from streaming data collection. The following can help with managing the data.

- Enriched, privacy-ready consumer and account profiles update automatically based on behavioural, transactional and operational data without manual data stitching.
- Data collected from across channels and systems is normalised into a standard taxonomy.
- Tag management and event forwarding manage data in real time.
- Data privacy and security is very important. Patented data governance framework ensures customer privacy and preferences, enables compliance with internal and external policies and gives role-based access to teams across the enterprise.
- Live preview of user's data helps businesses deliver latest recommendations.

And at the end we need a Holistic reporting which helps marketers to understand the impact of their campaigns so that they can create better experiences for their customers in the future.

The campaign reporting feature must facilitates the creation of dynamic reports to analyse the success of your campaigns in the real time. Depending on the complexity of your queries and calculations, the data must be aggregated into a list view or accessed in a format that makes it easy to generate marketing analytics reports.

I know all this is a lot to churn but great businesses go to any extent to provide great experiences to their customers. We at Adobe has all features mentioned above so that businesses can tailor their customer experience using Adobe experience cloud.

About the Author

Vikram Gupta a front-end architect at Adobe Systems with 12+ years of experience providing high-impact web solutions for diverse industry organizations. He loves working on improving web app performance and developer experience and sharing knowledge and giving back to the community. He likes traveling, reading books, and playing Badminton, and TT in his free time.

Kind Attention !

Prospective Contributors of Newsletter Bytes

Forthcoming Issues : Jauary 2023 : **Machine Learning for Big Data Analytics**

Please note that Cover Theme for the November 2022 issue is Machine Learning for Big Data Analytics. Please send your contributions by 15th December, 2022.

The articles should be authored as original text. Plagiarism is strictly prohibited.

Please note that Byte is a newsletter for members at large and not a research journal for publishing fullfledged research papers. Therefore, we expect articles written at the level of general audience of varied member categories. Equations and mathematical expressions within articles are not recommended and, if absolutely necessary, should be minimum. Number of pages of the articles should be at least 3 to 4 pages. Include a brief biography of four to six lines for each author with high resolution author photograph.

Please send your article in MS-Word format to Chief Editor, Dr. Tirimula Rao Benala in the email ids bytes.newsletter@gmail.com; Issued on the behalf of the Bytes Editorial Board.

Dr. Tirimula Rao Benala
Chief Editor

Coding to get into MAANG (Meta, Amazon, Apple, Netflix, Google)

Tharun Vambaravelli

Pursuing his master's in the stream of Computer science and Engineering at International Institute of Information Technology , Hyderabad. Email: tharunvambaravelli@gmail.com

Skills such as Problem-solving, expertise in solving algorithms and data structures, coding competition expertise, etc. are required for getting into MAANG.

Language and Fundamentals

Firstly choose a language and understand the basic fundamentals concepts such as variable declaration, basic syntax, data type and structures,flow control structures (conditions and loops), functional programming, recursion, object-oriented programming. Preferably, choose an object oriented language.

Time and Space Complexity

For a problem, there can be more than one solution. You need to come up with an optimal solution for a problem. If you come up with multiple solutions for a problem, how do you choose the optimal one? Here comes the concept of time and space complexity. It is really important to learn these concepts and given a solution you must be able to analyze the time and space complexity.

Data structures

Understanding data structures is integral to participating in competitive programming, as you will be faced with making decisions on what data structure to utilize to most efficiently solve the problem you are given. Some important Data structures are: Arrays, Stacks, Queues, Linked List, Binary Tree, Binary Search Tree, Graphs, Tries, Hash tables, Heap. Some of the advanced data structures such as : Disjoint sets, Fibonacci Heap, Segment tree, Fenwick tree, B tree, B + tree, Self balancing binary tree (RB, AVL).

Learning data structures alone is not sufficient but also trying to implement these data structures in your chosen language will be useful for better understanding. Try to understand and analyze the time and space complexities of the operations performed on each data structure. This will help you in choosing an appropriate data structure for solving a given coding problem.

Algorithms

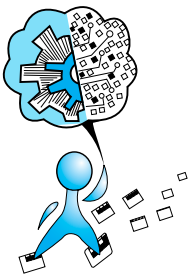
Algorithms are used to find the best possible way of solving a problem. Why to reinvent a wheel ? Algorithms can make life easier for developers by arming them with a range of general problem solving techniques. Choosing the standard algorithm is equally important as choosing an appropriate data structure for solving a problem. Some of the standard algorithms are: Searching & Sorting ,Greedy algorithms ,Dynamic Programming, Pattern Searching, Recursive and Backtracking algorithms, Geometric Algorithms, Graph Algorithms, Bit Algorithms, Number theory algorithms. Try to understand and analyze the time and space complexities for each algorithm. Give an attempt to implement each algorithm you have learnt in the chosen language.

Debugging

One common mistake a beginner makes is not using the debuggers. Try to use the debugger to debug the code and find out errors where you have made the mistake.

Practice

Try to solve at least 200 questions on any of the platforms like LeetCode, HackerRank, GeeksForGeeks, CodeChef, CodeForces etc. Try to stick to one or two coding platforms in the beginning phase until you are confident enough in solving questions and get into the habit of coding. Stick to learn topic wise (data structure and algorithms) in the early days of coding. Try to pick a data structure,understand and implement it , and solve problems on this chosen data structure. In a race of solving the problems, don't forget to build actually needed thought processes. It is more of quality rather than quantity while solving the problems. It is more important to solve problems from different topics.



Try to be consistent in solving problems i.e. try to solve at least 2 problems per day. Don't directly jump into solving the hard problems first as it may affect your confidence. Build confidence by first solving the easy problems,then medium and then hard problems.

Giving Contests and Competitions

After you think that you have got confidence in solving the problems, try to give contests in leetcode/codechef/ codeforces... Don't get demotivated if you can't solve a problem in the contest. Once the contest is finished, go through the discussion forums and try to understand various approaches. Even if you are able to solve the problem,try to check the other approaches which may be useful for solving other problems. Once you have enough practice in coding and giving contests regularly, you are now able to participate in competitions. Compete in competitions like ACM - ICPC, Google Code Jam, Facebook Hacker Cup. Happy coding :-)

About the Author

Tharun Sai is currently pursuing his master's in the stream of Computer science and Engineering at International Institute of Information Technology , Hyderabad.He is interested in Algorithms, Advanced Data Structures, Data systems,Operating systems.

Sails Software Solutions

We steer to coach, code, and conquer.

Sails Software Solutions is an ISO 27001 certified Company, which started its journey in May 2015 with a group of 7 passionate individuals and has now over 150 employees. Join our team of unmatched talent to learn and work on SaaS Development, Cloud Microservices, DevOps, Big Data, and Analytical Services and be a part of our coveted product development team.

Experience the thrill of working with our clients from the US and the UK, spread across Pharma, Banking, and HCM domains as we provide our expertise in developing Enterprise Development, Digital Transformation, Human Capital Management Solutions, and Cloud Transformation Solutions. Some of the niche technologies that we get our hands dirty with include Laravel, ReactJS, React-Native, Angular, .Net, Java, Elasticsearch, Mongo, Hadoop, Databricks, AWS, Azure, Docker, Kubernetes.

Life @ Sails

Work hard and play hard - we aim to go a long way with our crew while ensuring each one gets an opportunity to enjoy every moment spent here. Technical and soft skills training, CSR and other team building activities are some of the many ways that help us keep track of the pulse of all the employees regularly.

We welcome you to be a part of a family which embraces, trusts and recognizes every Sail'r.

As a proud partner to Microsoft Silver and Google Cloud, we aim to achieve much more in the near future.

Microsoft Silver Partner

Information Security Management ISO 27001 Certified

Google Cloud Partner

SAILS Software

+1-248-750-6252
+91 90303 51315
info@sailssoftware.com
www.sailssoftware.com

Offices: USA | Visakhapatnam | Hyderabad

India Headquarters:
Plot No. 12, Eshwar Sai Enclave On, 80 Feet Rd, Bakkannapalem, Madhurawada, Visakhapatnam, Andhra Pradesh - 530041