INTUIT

WiDS TLV

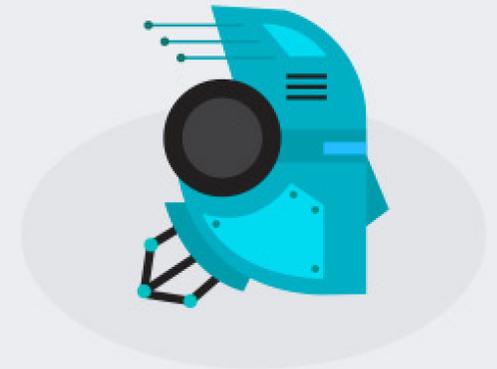# Women in Data Science Conference
Program

📅 Mon, March 5, 2018

🕐 09:00 AM - 03:30 PM

📍 Sheva, HaTsfira St 19 ,Tel Aviv-Yafo

Intuit

WiDS TLV

Round Table Topic	Facilitator	Bio
ElectroEuro	**Talia Kohen,** Founder and CEO at FemTech	Shortly after graduating from Cornell ECE in 2006, **Talia** went to work for Raytheon with a position focused on tracking and discrimination. She is currently completing a master's degree in computer science on the Evolution of Words in The Google Books Corpus.
Thinking Outside the Matrix: Are you a Data-Artist or Data-Mechanic?	**Nurit Vatnik,** Vice President of Data Science Excellence at SparkBeyond **Matar Heller**, Senior Data Scientist at SparkBeyond	**Ms. Vatnik** is Vice President of Data Science Excellence at SparkBeyond, where she leads a growing community of amazing data scientists (almost half of them women) who are steering the company's AI powered research platform. **Matar** is a senior data scientist at SparkBeyond, where she generates novel insights from data. Prior to SparkBeyond, she was a data scientist at Winton Capital and a fellow at Insight Data Science. Matar completed her PhD in neuroscience at the University of California at Berkele
Machine Learning without a Negative Set	**Dr. Einat Sitbon** Algorithm Developer at Evogene	**Dr. Sitbon** is working in the biotech industry for ten years, in various computational biology roles. In the past few years she is developing computational tools for analysis of biological data in Evogene, aimed at improving plants for agriculture. Dr, Sitbon holds Ph.d in Biology from Weizmann Institute of Science.
Challenges of Managing a Data Science Team	**Amit Weil**, Data Science Team Leader, Simplex	**Amit** was the first data scientist at Simplex, a Fintech company that provides worldwide fraudless payment processing. Amit created the very first machine learning models and today as data science team leader faces challenges of a rapidly growing business - both in company size and in data amounts.
Regaining Focus on Analytics	**Inbal Gilead,** Director of Data and BI, EverCompliant Ltd	**Inbal** is the Director of Data and BI at an Israeli startup that focuses of Data and Cyber Intelligence. Most intrigued by the ongoing Industry changes and with the huge potential data bring to organizations. Beginning her career 11 years ago as an Intelligence Analyst in the IDF where she founded the first BI team in the 8200 unit
Machine Learning - How Not to Lose Your Customer	**Nurit Peres**, Product manager, HARMAN Automotive Cybersecurity - TowerSec	**Nurit** is a product manager and UX lead for many years. In the last 5 years Nurit is involved in products that are heavily driven by Machine Learning. With a deep understanding of product management and User Experience Nurit approached Machine Learning with a fresh and unique perspective that have totally changes the course of these products.
When Can We Use Deep Learning Methods for Image Processing?	**Dr. Tamar Kashti**, Algorithm and Simulation Team Leader, Landa labs	As an Algorithm and Simulation team leader in Landa labs, **Dr. Kashti** is an Algorithm and Simulation team leader in Landa labs. Tamar is working on image processing algorithms and physical simulations in the last 8 years in Landa labs and HP Indigo. Tamar did her Ph.D. and M.Sc. in theoretical astrophysics and particle physics in Weizmann Institute of Science, where she did both analytical physics and physical simulations.
Measuring Your Model's Business Value	**Liat Friedman Antwarg,** Data Scientist at CBG	**Liat** holds a M.Sc in Information systems engineering from BGU, worked for few years in the industry in text analytics R&D team and currently a data scientist in the innovation labs at BGU. Liat also advise to companies interested in different types of data science solutions.
Defining the Prediction Model Target Variable – Challenges and Practical Solutions	**Einat Ben-Ari**, Director, Risk Analytics at PayPal	**Einat** is Director for Seller Risk Management at PayPal. She leads a global team of Data Scientists and Risk Analysts, responsible for managing losses for PayPal due to seller non-delivery and for managing seller funds availability policy and experience. Einat holds a B.Sc. in Industrial and Management Engineering and an M.Sc. in Applied Statistics.
Introducing Data Science into a Company	**Orian Sharoni**, Data Scientist at Soluto	**Orian** Works as a Data Scientist at Soluto, enthusiastic about using computational methods to better understand human behavior. Additionally, co-manages the Israeli branch of NeuroTech TLV, a non-profit organization aimed to bring together and create art and educational brain related projects. Loves to code interactive tech-art that changes people. Orian holds a Master degree in cognition from Tel Aviv University.
Technical Debt in Machine Learning	**Inbar Naor,** Data Scientist at Taboola	**Inbar** is a Data Scientist at Taboola, where she uses deep learning to develop recommendation systems at large scale. Before that she worked as a Data Scientist and as a software developer at several companies. She holds an MS.c in Computer Science with a focus on machine learning research, and a BS.c in Computer Science and Cognitive Science. On her spare time she is co-hosting Unsupervised, a podcast about data science in Israel, and organizing DataHack, a data-driven hackathon.
Privacy by Design in Autonomous Vehicle	**ADV Limor Shmerling Magazanik** Director of Strategic Alliances at The Privacy Protection Authority	**Limor** is the Director of Strategic Alliances at the Israeli Privacy Protection Authority, focusing on protecting the fundamental human right to privacy by regulation and enforcement in technology driven sectors. Limor is a privacy thought leader, responsible for policy delivery, promoting compliance and raising awareness to the importance of privacy in the age of the 4th industrial revolution - Data.

Lightning Talks Topic | Facilitator | Bio

Lightning Talks Topic	Facilitator	Bio
Training Word2vec for Learning Word Relationships for a Specific Domain	**Efrat Blaier,** Co-Founder and CTO of mmuze	**Efrat** Blaier is the Co-Founder and CTO of mmuze for the past 4 years. She have 12 years of experience in the fields of Machine Learning and Natural Language Processing. After completing her BS.c in computer science on the president's list at the Technion at the age of 20, she served in the army for 6 years leading the textual search engine for the entire intelligence unit. Later in her career at Kontera as one of the team leads on machine learning for a big data advertising system . In 2014, Efrat joined 2 of my co-founders to start a compan y - mmuze and serve as its CTO. Today we are 9 employees and growing.
Finding A Needle in a Haystack - Anomaly Detection in Time-Series Data	**Shir Meir Lador,** Data Science Team Leader at Intuit	**Shir** is data science team lead at Intuit, the global leader in financial management software. She is the co-founder and organizer of PyData Tel Aviv meetups and a co-host of the podcast "Unsupervised" - A Podcast about Data Science in Israel. Shir is a well-known speaker at various machine learning and data science conferences and meetups. Shir holds an MSc in electrical engineering and computers with a major in machine learning and signal processing from Ben Gurion University.
ML for Preprocessing Prediction of Medical Imaging Algorithms.	**Bella Fadida Specktor**, Sr. Algorithms Developer at Philips	**Bella** Fadida Specktor is a Sr. Algorithms Developer at Philips working on Machine Learning and Deep Learning problems in the Medical Imaging domain. She holds MSc degree in Computer Science from Haifa University under the supervision of Dr. Margarita Osadchy and Prof. Daniel Keren. She is also the organizer of Haifa Machine Learning meetup
Beyond What is Visible to the Human Eye; Leveraging Machine Learning to Predict Asthma/ COPD Exacerbation.	**Daphna Laifenfeld,** Senior Director, Head of Personalized Medicine at Teva	**Dr. Daphna** Laifenfeld is Head of Personalized Medicine and Diagnostics (PMD), at Teva's Global R&D. Daphna has dedicated her career to the field of personalized and translational medicine, with a focus on genomics and biomarker implementation through to diagnostic development. Before joining Teva, Dr. Laifenfeld engaged in translational/biomarker development activities within the pipeline of multiple top-10 pharma, including Roche, GSK, Pfizer, as well as consumer product companies. In academia, Dr. Laifenfeld studied molecular pathways and pharmacogenomics in CNS, focusing on the biology underlying Alzheimer's disease at Harvard University, and on pharmacogenomics of Major Depression therapeutic response at the Technion. Dr. Laifenfeld received her PhD from the Technion – Israel Institute of Technology.
Surfacing Hidden User Data	**Noa Haas,** Data Scientist at Intuit	**Noa** works as a Data Scientist at Intuit. In her work, she gets to explore and apply advanced learning methodologies, creating real-life benefit for Intuit's customers. She is passionate about non-standard data-related problems, and enjoys the process of exploring and solving such riddles. Before working for Intuit, Noa worked at the Tel Aviv University Statistics Lab. Noa is currently pursuing her MSc in Applied Statistics.
RNN for Session Based Recommendation	**Asi Messica,** Lecturer at Ben-Gurion University of the Negev	**Asi** is a PhD student in the Department of Software and Information Systems Engineering at Ben-Gurion University of the Negev, under the supervision of Prof. Lior Rokach. Prior to starting her PhD, Asi held several senior Product Management, Product Marketing, and Engineering Management positions in leading software companies such as EMC and SAP.
The Automatic Poet - Using Deep Networks for Text Generation	**Sigal Shaked,** PhD Student and Lecturer, Ben Gurion University.	**Sigal** has a seven years' experience as a Data Scientist in the security community, and now is on her fourth-year doctoral student in the department of Software and Information Systems Engineering at Ben Gurion University. Throughout her doctoral studies, she developed a course that provides tools for data scientists. Sigal also specializes in deep learning techniques through their application in Natural Language Processing and data privacy tasks.
New Methods for Sequential Design of Sensitivity Experiments	**Rotem Rozenblum,** Data Scientist at Intuit	After completing her B.Sc. in mathematics and statistics and M.Sc. in applied statistics from TAU, Rotem has been working as a research data scientist for various fields of industry. Rotem is a passionate mathematician and statistics researcher. She thinks of statistics as the way to see how mathematical principles are manifested in real-life. Rotem's greatest interest is in probability theory and in Bayesian experiments design in particular.

INTUIT

WiDSTLV

Lightning Talks Topic | ## Facilitator | ## Bio

Lightning Talks Topic	Facilitator	Bio
Offline Solutions to Online Problems	**Dalya Gartzman**, Algorithm Researcher at Via - On-Demand Transit	**Dalya** is mathematician at heart, an algorithmatician in practice. After receiving her MSc. in math from TAU, she ventured into data science and algorithms domains, glad to discover that the real world is at least as exciting as the abstract. Currently algorithms researcher at Via - On Demand Transit, where she apply both of her theoretical and practical passions.
The labels are Out There: Creative Ways to Obtain Labeled Data	**Lotem Peled**, Data Scientist and NLP Researcher at Fiverr	**Lotem** is a Technion graduate with a master's degree in NLP (a student of Roi Reichart), working as a data scientist and researcher at Fiverr, a startup in Hertzelia.
Implementing Machine Learning Methods for Identifying Trends and Hidden Connections in the Dark Net	**Sigalit Bechler** Senior Data scientist at Intuit	**Sigalit**, passionate and fascinated with science and the understanding of nature and its governing laws. Sigalit has a diverse academic background; B.Sc's in electrical engineering and in physics, M.Sc in Physics, and two years of Ph.D. research in bioinformatics. Sigalit has been working over the past few years as a researcher, data scientist and algorithm developer in several of Israel's market leading hi-tech companies

Tutorial

Tutorial	Facilitator	Bio
NLP Tutorial	**Hila Gonen**, PhD Student at Bar-Ilan University in the Field of NLP and Deep Learning	**Hila** is a PhD student at Bar-Ilan university in the field of NLP and deep learning, under the supervision of Dr. Yoav Goldberg. She is fascinated by languages and is interested in relations between different languages and the way multilingual signals can be used for various tasks. Lately, she has been focusing also on research in the area of code-switching.

ElectroEuro

**Talia Kohen,
Founder and CEO
at FemTech**

Abstract

We designed a decentralized virtual currency known as the ElectroEuro for trading energies through an energy bank in Europe, driving a low carbon economy. The use of the greener energies would promote decarbonization, and the monetization would make it accessible and practical. The concept was to unite Europe through electricity like the Euro. This currency is similar to BitCoin in that it is universal and there is a finite quantity of it. The transaction of energy is carried out through it, and it can be bought through goods that do not promote carbonization.

Points for Discussion:

1. Why is the challenge of electrification so important?
2. Crypto-currencies and ones that are specialized
3. What technology is important for making a startup that focuses on this?

Bio:

Shortly after graduating from Cornell ECE in 2006, Kohen went to work for Raytheon with a position focused on tracking and discrimination. She is currently completing a master's degree in computer science on the Evolution of Words in The Google Books Corpus. She has implemented three algorithms, two known, and one she designed and later found in a textbook, and with those has been able to analyze one billion lines of text in two-and-a-half hours. During her master's work, Kohen independently developed by mere visualization a result in mathematics known as "Euler's F-Vector".

Kohen has received numerous awards and honors including: Anita Borg Birthday Celebration Director in Israel; a Google OutStander; Google Anita Borg Scholarship for Women in Europe and the Middle East Finalist; Google Campus Ambassador; Microsoft Israel Women of Excellence Program; Microsoft Excellence Summer Camp; ACM XRDS feature issue editor for the IoT Edition; IEEE International Radar Conference Poster Session Co-Chair and Steering Committee Raytheon Individual Performer Achievement Award – Ionospheric study Raytheon Spot Award; Raytheon Women's Network MDC Site Representative; and was a Delegate to Grace Hopper Conference for Women In Computing. Kohen is the CEO of FemTech, a community for women in STEM in Israel. Upon completion of her masters degree, Kohen plans to earn a

INTUIT.

WIDSTLV

Thinking Outside the Matrix: Are You a Data-Artist or Data-Mechanic?

**Nurit Vatnik,
Vice President of Data Science Excellence at SparkBeyond &**

**Matar Heller,
Senior Data Scientist at SparkBeyond**

Abstract

What is the first thing you do when tackled with a new challenge?

Do you ever stop to imagine what the best possible insights might look like? What inputs would lead you to their discovery? Weave a story around them?

We argue that this creative part of the data science process is the most critical for success - beginning with the end in mind, encompassing the wider complexities of the problem domain, creatively imagining a desired outcome - and only adhering to inevitable limitations after exploring the end goal and the story behind it.

We demonstrate through real word case studies how widening the research scope is critical to breakthrough discoveries. We share best practices for implementing a creative data science process, ideas for useful data sources, and a framework for business-driven hypothesis search that enables data practitioners to think outside the matrix.

Points for Discussion

Intro: explain methodology guidelines
Guided Discussion - We will scope together a real use case:

- Problem definition (what is our goal?)
- What could be potential target variables?
- How would you define the sample population?
- Pause to imagine the best insights you expect to find
- What factors might influence the target?
- Which datasets could we use?
- Which transformations should we use to engineer features?
- Share & discuss actual results
- If time allows - explore other use cases that you bring to the table
- Wrap up - summary & best practices

Bio

Ms. Vatnik is Vice President of Data Science Excellence at SparkBeyond, where she leads a growing community of amazing data scientists (almost half of them women) who are steering the company's AI powered research platform. She is passionate about data driven storytelling and the art of data science. Before joining the founding members of Sparkbeyond Nurit was an active edu-preneur leveraging technology to democratize education at MIndset incubator & Education Cities foundation . Prior to that Nurit established and lead the advanced analytics line-of-business at Amdocs, serving clients like AT&T, Sprint, Vodafone & Singtel - helping them translate operational data into actionable knowledge. In her academic research at Tel Aviv University, Nurit focused on Deep Learning, employing ANNs to Financial Forecasting. She is a KDD cup award winner, a mother of two teenage boys, and an avid Jazz fan.

Matar is a senior data scientist at SparkBeyond, where she generates novel insights from data. Prior to SparkBeyond, she was a data scientist at Winton Capital and a fellow at Insight Data Science. Matar completed her PhD in neuroscience at the University of California at Berkeley, where she tracked the progress of a thought in the brain by recording activity from surgically implanted electrodes. Before her PhD, Matar was a preschool teacher and a tank instructor (though not at the same time). She is passionate about expanding opportunities for women in STEM fields and is a mother to a precocious toddler who surprises her every day.

Machine Learning Without a Negative Set

Dr. Einat Sitbon, Algorithm Developer at Evogene.

INTUIT

WiDSTLV

Abstract

For classification models, often a positive class for the training set is available, while no negative set can be easily devised. This problem commonplace when data is derived from experiments published publicly, where negative results are not documented. It could also occur where there is no clear definition to the negative term. This problem has a high impact when the prevalence of the negative class is assumed to be high. There are several approaches to such "PU learning" (positive-unlabeled learning). One approach assumes that the positive and unlabeled samples provide sufficient information for learning. I will describe how we used this approach, its strengths and its weaknesses. The participants will discuss various approaches, and their relevance in their field of work.

Bio

I am working in the biotech industry for ten years, in various computational biology roles. In the past few years I develop computational tools for analysis of biological data in Evogene, aimed at improving plants for agriculture. This includes machine learning for the identification of insecticide genes, and phenotypic and genotypic analysis of complex traits in plants. In my work, I use various data sources, ranging from genomic information, to plant phenotype measurement. I am currently using machine learning algorithms and exploring deep learning, in addition to using more traditional modeling. During my studies at Weizmann Institute towards a PhD, I used various computational tools to explore biological questions. I am fascinated by the complexity and variability of biology, and the intriguing questions and challenges it presents.

Points for Discussion

- set is unavailable
- Approaches used by each participants
- Pros and cons for different approaches
- How to avoid over-fitting pitfalls
- Risk estimation methodologies

Challenges of Managing a Data Science Team

Amit Weil,
Data Science Team Leader, Simplex

INTUIT

WiDSTLV

Abstract

Data science is a relatively new field for many companies. Managing data science teams raises key challenges, this session will try to provide different perspective and opinions on some of them.

Points for Discussion

- Research tasks management
  - How can tasks time be estimated?
  - How to tell that a research is sufficient or doomed?
  - If and how to divide research tasks between different individuals?
- Using R&D tools and methods within data science teams
  - What sets of standard tools and methods should be leveraged?
  - Should data science team work in an agile methodology?
  - How to manage experiments, datasets and reports?
- Data science as part of the bigger organization
  - How does data science teams fit within the organization?
  - What kind of problems should the data science team solve? are there other alternatives?
  - How to plan research in a dynamic business environment?
  - Best practices in cooperating with other R&D teams
  - How to integrate research teams somewhat flexible timelines in the bigger company Gantt?

Bio

Amit was the first data scientist at Simplex, a fintech company that provides worldwide fraudless payment processing. Amit created the very first machine learning models and today as data science team leader faces challenges of a rapidly growing business - both in company size and in data amounts.

Regaining Focus on Analytics

**Inbal Gilead,
Director of Data and BI, EverCompliant Ltd**

INTUIT

WiDS TLV

Abstract

It's tempting to think that our organizations are struggling through a lack of data. The reality is, that most of us have more than enough to make insightful decisions but the data we work so hard for, is poorly managed and very often misused.

Companies invest all their resources in collecting the data and not enough in analyzing and understanding what to do with it. Dealing with issues such as unstructured entities extraction, classification and deep learning can be both challenging and rewarding for a company. However, understanding and utilizing the results are the most important parts.

In order to generate the real value from the vast data being collected, organizations should learn how to properly manage their data and know how to utilize it to its full potential.

Bio

I began my career 11 years ago as an intelligence analyst in the IDF where I founded the first BI team in the 8200 unit. It was clear to me that this was the field I was passionate about and have been hooked ever since. I've been an analyst of all types, a BI developer, analytics expert, data product manager. Currently, I am a Director of Data and BI at an Israeli startup that focuses of Data and Cyber Intelligence. Most intrigued by the ongoing Industry changes and with the huge potential data bring to organizations.

Points for Discussion

1. How data should be managed in an organization in order to enable strategic utilization
   - Where in the organizational structure this should be managed.
   - Which data functions are required.
   - Analyze mistakes companies are making in this respect.
2. How can data leadership drive senior-management to be more involved and make better decisions
   - · Discuss the importance of a «Data Strategy»
   - · Ways for measuring data impact

INTUIT

WiDSTLV

Machine Learning - How Not to Lose Your Customer

Nurit Peres,
Product manager,
HARMAN Automotive
Cybersecurity -
TowerSec

Abstract

When developing an application that is based on Machine learning, it is easy to get caught up with algorithms and data issues and lose sight of the user. Better knowledge of users can influence our data selection and algorithms. In addition, investing in user trust may improve not only user loyalty but also eventually the algorithm itself.

Points for Discussion

• Why and how user goals influence my algorithm?
• How do they influence the data?
• Why is user context important and when should I consider it?
• Possible ways to gain user trust.
• Why is it important to give information to users about how my algorithm work?
• How does giving such information may influence my code?
• Why is it important to be transparent about the way the algorithm works?
• And how does that improve the code?

Bio

Nurit Peres is a product manager and UX lead for many years, with experience on a variety of products and platforms. In the last 5 years Nurit is involved in products that are heavily driven by Machine Learning. With a deep understanding of product management and User Experience Nurit approached Machine learning with a fresh and unique perspective that have totally changes the course of these products.
Nurit is also a Google Expert and a mentor in Google Launchpad.

When Can We Use Deep Learning Methods for Image Processing?

Dr. Tamar Kashti, Algorithm and Simulation Team Leader, Landa

INTUIT

WiDSTLV

Abstract

Classically, image processing and machine learning have distinct roles. While image processing mainly focused on processing raw images and preparing them for other tasks (e.g. – edge detection), while computer vision deals with understanding the meaning of those images (e.g. – classification). Recently, deep learning methods bypassed all other machine learning methods in solving classification problems, and became the main tool for machine learning and artificial intelligence. Resent papers show that deep learning can also solve classical image processing, e.g. super-resolution, edge detection. When can we use deep learning methods for image processing?

Points for Discussion

A list of points for discussion:
- Image processing vs. machine learning vs. deep learning.
- Examples for classical image processing problems that were solved using deep learning.
- When can we use deep learning methods for image processing?

Bio

Tamar Kashti is an Algorithm and Simulation team leader in Landa labs. Tamar is working on image processing algorithms and physical simulations in the last 8 years in Landa labs and HP Indigo. Tamar did her Ph.D. and M.Sc. in theoretical astrophysics and particle physics in Weizmann Institute of Science, where she did both analytical physics and physical simulations.

Measuring Your Model's Business Value

Liat Friedman Antwarg,
Data Scientist at CBG

INTUIT.

WiDSTLV

Abstract

A very important aspect in Data science is how and what to measure. How to reflect in the best way the value that the machine or deep learning algorithm gives. What will cause the managers to say - this is exactly what I was looking for.

We will discuss how the measures are related to the purpose of the problem we are trying to solve. For example, for marketing purposes do we really need to know the accuracy of the algorithm for all the classes? Or is it enough to be very accurate in predicting a specific class which the marketing department is interested in?

We will discuss what should be measured and how to present the results, how to translate the measures to working hours of employees or money saved or higher revenues to the company.

Bio

Liat Friedman Antwarg, have M.Sc in Information systems engineering from BGU, worked for few years in the industry in text analytics R&D team and currently a data scientist in the innovation labs at BGU, starting my Ph.D soon. I also advise to companies interested in different types of data science solutions.

I am interested in machine learning, recommender systems, NLP and personalization and especially in the connection between machine learning and business cases.

Every company I hear about, I think instantly what can be done with their data, which problem it can solve and how and what can be improved.

Points for Discussion

Measuring from a business point of view
• How and what to measure
• Cost and benefit of false and true decisions
• The relation between a data scientist and the business units

Defining the Prediction Model Target Variable – Challenges and Practical Solutions

Einat Ben-Ari, Director, Risk Analytics at PayPal

INTUIT®

WiDS TLV

Abstract

When developing a prediction model, defining the target variable might seem like a relatively straightforward task but it often turns out to be complex and challenging. It is very important to align the model target variable with the business problem as well as with how the model prediction will be used. There is often a gap between the optimal target variable and the one we end up developing due to various constraints such as data availability or reliability, data complexity, and time and resource constraints. In this round table we will share challenges in defining target variables and the approaches taken to tackle those challenges. The domain we shall focus is prediction models for Payments Risk Management but the challenges and some of the practical solutions can be applicable to other domains as well.

Bio

Einat Ben-Ari is Director for Seller Risk Management at PayPal. She leads a global team of Data Scientists and Risk Analysts, responsible for managing losses for PayPal due to seller non-delivery and for managing seller funds availability policy and experience. Einat has been working for PayPal since 2010. In her prior role at PayPal she led the EMEA Advanced Risk Sciences team, which developed customized risk solutions, mitigating sophisticated fraud trends while supporting the unique business goals of the region. Prior to PayPal, Einat held several management positions at eBay in the Trust and Safety and Business Analytics departments. Before that she worked at Amdocs, focusing on developing Data Science software and solutions for Communications companies.

Points for Discussion

• Scenarios in which the value of the target variable for historical cases is unknown
• Defining logic to determine the value of the target variable
• Optimizing precision vs. recall
• Applying manual labeling
• Leveraging Active Learning for labeling

Introducing Data Science Into a Company

Orian Sharoni,
Data Scientist at Soluto

INTUIT.

WiDSTLV

Abstract

Data science is more than a set of tools ready to be used at a given task; data science is a way of thinking. Integrating data processes that drive a product forward require patience and planning, as there are numerous challenges along the way. The table discussion would be devoted to companies that want to add data scientists or that have started to do so recently.

Points for Discussion

- What is the best way to add a data team to the production process?
- How to create an enthusiastic atmosphere among the company's personnel (programmers, product managers, designers and managers alike) regarding statistics? #Are these questions related to product management or data science?

Bio

Orian Sharoni Works as a data scientist at Soluto, enthusiastic about using computational methods to better understand human behavior. Additionally, co-manages the Israeli branch of NeuroTech TLV, a non-profit organization aimed to bring together and create art and educational brain related projects. Loves to code interactive tech-art that changes people. Orian holds a Master degree in cognition from Tel Aviv University.

INTUIT

WiDSTLV

Technical Debt in Machine Learning

Inbar Naor,
Data scientist at Taboola

Abstract

In this discussion I will present ideas from different papers on the subject and lessons learned from my work at Taboola, where 30 Data Scientists collaborate to build systems that serves billions of requests a day.

We will explore the different causes for technical debt in ML and use the joint experience of participants to discuss possible solutions.

Bio

Inbar is a Data Scientist at Taboola, where she uses deep learning to develop recommendation systems at large scale. Before that she worked as a data scientist and as a software developer at several companies. She holds an MS.c in Computer Science with a focus on machine learning research, and a BS.c in Computer Science and Cognitive Science.

On her spare time she is co-hosting Unsupervised, a podcast about data science in Israel, and organizing DataHack, a data-driven hackathon.

Points for Discussion

- Different types of technical debt (e.g - boundary erosion, hidden feedback loop, data dependencies, system's reaction to changes in the real world, etc...)
- Possible solutions for those problems
- Best practices for developing ML systems
- Challenges in moving research code to production

Privacy by Design in Autonomous Vehicles

Limor Shmerling Magazanik,
Director of Strategic Alliances at The Privacy Protection Authority

INTUIT®

WiDSTLV

Abstract

"By 2020 autonomous vehicles will generate about 4,000 gigabytes of data a day (Forbes). The automated vehicles industry is expected to create a global market worth £900bn by 2025 (Catapult.org. uk). The data generated may include highly personal and sensitive information, such as geolocation, contents of communications via mobile phones, passengers' speech and actions via sensors and microphones and more. The digital interfaces may expose the vehicle to cyber attacks that may have safety implications, and also risks to data privacy. The stakeholders will grow in numbers and originate from various countries. The complexity of the algorithms managing the vehicle will grow and likely enter the realm of artificial intelligence. The possible benefits to human society are vast and varied. A decrease in accidents and injuries due to human error, which constitute 90% of collisions; improved traffic flow, energy saving and reduction in air pollution; enhanced mobility for people with disabilities, aging populations, people who cannot afford a car or do not wish to drive. At the same time this evolution will present new challenges to our diminishing privacy.

Points for Discussion
Some of the questions asked in this panel will include: what are the various challenges to privacy in the autonomous vehicles ecosystem? What are the principles of Privacy be Design? How can they offer mitigation? How can developers and manufacturers in the Auto industry implement them? What is the role of Industry Standards?

Bio
Adv. Limor Shmerling Magazanik (CIPP/E, CIPP/US, CIPM) is Director of Strategic Alliances at the Israeli Privacy Protection Authority, focusing on protecting the fundamental human right to privacy by regulation and enforcement in technology driven sectors. Limor is a privacy thought leader, responsible for policy delivery, promoting compliance and raising awareness to the importance of privacy in the age of the 4th industrial revolution - Data. She is a lecturer on Data Protection at the IDC Herzliya, Redziner School of Law, and holds an L.L.B. and L.L.M in Law, and an M.A in Literature, all from Tel-Aviv University. Her background includes positions in the private sector, in law firms and the high tech sector.

Training Word2vec for Learning Word Relationships for a Specific Domain

**Efrat Blaier,
Co-Founder and
CTO of mmuze**

INTUIT.

WiDSTLV

Abstract

Word2Vec has been around for a few years and shows great promise in exploring word relationships. Both Google and Facebook released English models that allow a researcher to explore those in the English language. However, training word2vec for a specific context (domain) is beneficial as well, as it will allow to uncover special word relationships for this specific domain. In this talk I will show how we trained this for the Fashion domain - which data and software we used and common errors a researcher may encounter.

Bio

My name is Efrat Blaier and I am the co-founder and CTO of mmuze for the past 4 years. I have 12 years of experience in the fields of Machine Learning and Natural Language Processing. After completing my first computer science degree on the president's list at the Technion at the age of 20, I served in the army for 6 years - my last role there was leading the textual search engine for the entire intelligence unit. After that I have worked at Kontera as one of the team leads on machine learning for a big data advertising system (a company acquired by Amobee). In 2014, I joined 2 of my co-founders to start a company - mmuze and serve as its CTO. Today we are 9 employees and growing. My presentation is around a specific problem we are trying to solve of understanding a consumer talking about products and creating an AI agent that mimics the behavior of a sales representative for our customers (e-commerce stores).

INTUIT®

WiDSTLV

Finding A Needle in a Haystack - Anomaly Detection in Time-Series Data

Shir Meir Lador, Data Science Team Leader at Intuit

Abstract

How quickly can you detect anomalous behavior within your systems? To solve this common problem, we are building a meta-model which rapidly detects anomalies in different types of system logs. With automatic feature and model selection, we are able to flexibly account for a wide variety of behaviors in time series data, both those seen in the development stage and new challenges.

Bio

Shir is data science team lead at Intuit, the global leader in financial management software. She is the co-founder and organizer of PyData Tel Aviv meetups and a co-host of the podcast "Unsupervised" - A Podcast about Data Science in Israel. Shir is a well-known speaker at various machine learning and data science conferences and meetups. Shir holds an MSc in electrical engineering and computers with a major in machine learning and signal processing from Ben Gurion University.

INTUIT.

ML for Preprocessing Prediction of Medical Imaging Algorithms,

Bella Fadida Specktor,
Sr. Algorithms Developer at Philips

Abstract

Advanced medical imaging algorithms can provide extremely valuable information to the radiologists, but they might sometimes be very time consuming. Being able to run the algorithms in advance can be a possible solution. However, we do not know which algorithm to run on a given dataset. In this talk I will present a dynamic machine learning solution for predicting which advanced visualization (AV) algorithm needs to be applied on a given series using textual metadata of the scan. It incorporates a Bag of Words (BOW) feature extractor and a Random Forest classifier. The approach was tested on two datasets from clinical sites which use different languages and varying scanner models. We show that even without feature extraction, sensitivity of above 90% can be reached on both of them. By using BOW feature extractor, precision and sensitivity can usually be further improved. Using such a system will ultimately minimize the time to launch studies and improve radiologists reading time efficiency.

Bio

Bella Fadida Specktor is a Sr. Algorithms Developer at Philips working on Machine Learning and Deep Learning problems in the Medical Imaging domain. She holds MSc degree in Computer Science from Haifa University under the supervision of Dr. Margarita Osadchy and Prof. Daniel Keren. She is also the organizer of Haifa Machine Learning meetup.

INTUIT.

WiDSTLV

**Beyond What
is Visible to the
Human Eye;
Leveraging
Machine Learning
to Predict
Asthma/COPD
Exacerbation**

**Daphna
Laifenfeld,
Senior Director,
Head of
Personalized
Medicine at Teva**

Abstract

This talk will concentrate on the digital health solutions in the respiratory area in Teva. Asthma and COPD continue to present healthcare challenges, with many patients remaining poorly controlled. Early identification of personalized risk for asthma/COPD exacerbations will enable targeted intervention programs for this patient population.

I will present how we leverage Machine Learning to assess the risk of an asthmatic patient to experience asthma exacerbation(s) in the future; Advanced analytical methods are being used to build several predictive models, using data from smart medical devices, data from wearables, as well as patient-specific data from Electronic Medical Records (EMRs). I will also briefly show analysis of Time Series data by Deep Learning methods in order to detect sleep/awake patterns.

Bio

Dr. Daphna Laifenfeld is Head of Personalized Medicine and Diagnostics (PMD), at Teva's Global R&D.

Daphna has dedicated her career to the field of personalized and translational medicine, with a focus on genomics and biomarker implementation through to diagnostic development. Before joining Teva, Dr. Laifenfeld engaged in translational/biomarker development activities within the pipeline of multiple top-10 pharma, including Roche, GSK, Pfizer, as well as consumer product companies. In academia, Dr. Laifenfeld studied molecular pathways and pharmacogenomics in CNS, focusing on the biology underlying Alzheimer's disease at Harvard University, and on pharmacogenomics of Major Depression therapeutic response at the Technion.

Dr. Laifenfeld received her PhD from the Technion – Israel Institute of Technology.Sigalit Bechler, passionate and fascinated with science and the understanding of nature and its governing laws. Sigalit has a diverse academic background; B.Sc's in electrical engineering and in physics, M.Sc in Physics, and two years of Ph.D. research in bioinformatics, after which she decided to join the fascinating world of data science in Israel's hi-tech industry. Sigalit believes in the power of the infusion of ideas and techniques from different fields, and has been working over the past few years as a researcher, data scientist and algorithm developer in several of Israel's market leading hi-tech companies.

Surfacing Hidden User Data

**Noa Haas,
Data Scientist
at Intuit**

Abstract

Mint is a personal assistance app for financial management. In Mint, user's financial data is being aggregated to provide a smart and useful financial management. The app collects the data from the various accounts the user may have, and it is accessible because the user informs Mint the existence of these accounts. It is common, however, for Mint users to forget to introduce the app with all of their accounts - after all, the need for Mint arise from the (common) reality of owning multiple accounts in multiple financial institutes. Once the provided data is incomplete, and Mint doesn't have a full picture of the financial condition of the user, any analysis, predictions and recommendations provided to the user may not be accurate. Being a smart advisor, it is crucial for Mint having a complete picture of the users financials. We set for ourselves the goal of tracing these users with unlisted accounts - using the data they do provide, in an unsupervised framework. Moreover, we try to predict the bank identity of those accounts, so that the road to adding them will be as short and smooth as possible for the users. In this talk I will describe our methodology in solving this two-components task, by mainly leveraging the textual data we see for the financial transactions' descriptions in the existing data, as well as our unique domain's characteristics.

Bio

Noa works as a Data Scientist at Intuit. In her work, she gets to explore and apply advanced learning methodologies, creating real-life benefit for Intuit's customers. She is passionate about non-standard data-related problems, and enjoys the process of exploring and solving such riddles. Before working for Intuit, Noa worked at the Tel Aviv University Statistics Lab. Noa is currently pursuing her MSc in Applied Statistics. learning, under the supervision of Dr. Yoav Goldberg. She is fascinated by languages and is interested in relations between different languages and the way multilingual signals can be used for various tasks. Lately, she has been focusing also on research in the area of code-switching.

INTUIT

WiDSTLV

RNN for Session Based Recommendation

Asi Messica,
Lecturer at Ben-Gurion University of the Negev

Abstract

A session-based recommendation where user is anonymous and the recommendations are based on short session available data instead of long user history is relatively hardly dealt in the recommender systems community. However, anonymous user recommendation is common in real-life for occasional user scenario when the user is not logged in or the user-id is not tracked for either technical or privacy reasons. Recurrent neural network (RNN) that preserves the order of user's clicks outperforms the item-to-item similarity algorithm which is considered state of the art for this task (recall@10 value of 42% based on first three clicks, versus 28% by item to item similarity).

In this talk, I'll present how to process user's clickstream in real time to provide effective recommendations using RNN. A detailed description of the data pre-processing, neural network architecture, tuning best practices and experiment results will be provided. In addition, a link to implementation using public dataset to provide a starting point for further experimentation will be provided.

Bio

Asi Messica is a PhD student in the Department of Software and Information Systems Engineering at Ben-Gurion University of the Negev, under the supervision of Prof. Lior Rokach.
Asi's PhD research is focused in incorporating dynamic pricing and personal discounts into
recommender systems to further improve its business effectiveness. Prior to starting her PhD,
Asi held several senior Product Management, Product Marketing, and Engineering Management
positions in leading software companies such as EMC and SAP.

INTUIT

WiDSTLV

The Automatic Poet - Using Deep Networks for Text Generation

Sigal Shaked,
PhD Student and Lecturer, Ben Gurion University.

Abstract

The generation of meaningful sentences is central to many natural language processing applications. In this lightning talk, we will discuss the use of deep network-based models for text generation. We will examine the generator's performance on various types of texts and demonstrate that the generated text preserves characteristics of the original text.

Bio

My name is Sigal Shaked and I am a fourth-year doctoral student in the department of Software and Information Systems Engineering at Ben Gurion University. My research deals with data generation as a method for privacy preservation. Throughout my doctoral studies, I developed a course that provides tools for data scientists, I taught the course five times, at Ben-Gurion University, Shenkar College, and Bank Leumi. During this period, I also specialized in deep learning techniques through their aplication in Natural Language Processing and data privacy tasks.
Before my doctoral studies, I served as a data scientist in the security community for about seven years, while working on a wide variety of data domains.

New Methods for Sequential Design of Sensitivity Experiments

Rotem Rozenblum, Data Scientist at Intuit

INTUIT.

WiDSTLV

Abstract

This work proposes an algorithm for a sensitivity testing experiment protocol. In such experiments the researchers are often interested in testing manufacturer's statements or setting thresholds for safe use of products. For instance, an experiment that tests the tendency of a device which contains some explosive materials, to detonate when electrical current is passed through its system. Sensitivity experiments are usually characterized by a controlled stress variable and a binary outcome. In this case the electrical current is considered as "stress" that is being applied, and the detonation is an event that either occurs or doesn't. This problem can also be viewed as a version of the continuous multi-armed bandit problem. As for now, current existing method aim to construct designs that produce good maximum likelihood estimates, which, in small samples are known to perform poorly. In this talk I will discuss a sequential approach, with the purpose to construct an entropy-based design to dynamically shape the collection of data, while working in a Bayesian framework for a more accurate parameter estimation.

Bio

After completing her B.Sc. in mathematics and statistics and M.Sc. in applied statistics from TAU, Rotem has been working as a research data scientist for various fields of industry. Rotem is a passionate mathematician and statistics researcher. She thinks of statistics as the way to see how mathematical principles are manifested in real-life. Rotem's greatest interest is in probability theory and in Bayesian experiments design in particular.

INTUIT

WiDS TLV

Offline Solutions to Online Problems

Dalya Gartzman,
Algorithm Researcher at Via - On-Demand Transit

Abstract

We live in an online world. As consumers we expect our demands to be answered on the spot, and as creators we design our products to answer consumers' immediate requests. In Via (https://ridewithvia.com/), for example, the minute you ask for a ride, you will be assigned a van - a live imitation of a greedy algorithm. But what if the algorithm had the input that in a minute from now I will also ask for a ride with similar origin and destination? then it will be able to optimize the assignment - it will keep you waiting for just a minute, so that we can ride together, even more cheaply and economically than ever!
To address this issue, we will talk about a project I lead in Via, where the goal was to imagine we knew all requests in advance, and make the best assignment to vans based on this "omniscience". In this talk I will outline our motivation for this project, and present some benefits that we gained from it, in hope to inspire other researchers to venture into the offline solutions domain.

Bio

A mathematician at heart, an algorithmatician in practice. After receiving my MSc. in math from TAU I ventured into data science and algorithms domains, glad to discover that the real world is at least as exciting as the abstract. Currently algorithms researcher at Via - On Demand Transit, where I apply both my theoretical and practical passions.

10

The Labels are Out There: Creative Ways to Obtain Labeled Data,

Lotem Peled,
Data Scientist and NLP Researcher at Fiverr

INTUIT

WiDSTLV

Abstract

What should we do if we come up with a great new task, or a new look on some problem, but there just isn't any pre-existing labeled data set? Moreover, what do we do if our task is too vague for crowd-sourced labeling (for examples, tasks that involve sentiment, humor, etc...)?
In this talk I will present two creative ways in which I've approached this issue, for NLP tasks such as Sarcasm Interpretation and Punctuation Prediction.

Bio

Lotem is a Technion graduate with a master's degree in NLP (a student of Roi Reichart), working as a data scientist and researcher at Fiverr.

INTUIT®

WiDS TLV

Implementing Machine Learning Methods for Identifying Trends and Hidden Connections in the Dark Net

Sigalit Bechler,
Senior Data Scientist at Intuit

Abstract

The TOR network, also known as the Dark Net, is not the typical place for data mining. Since the Dark Net is not indexed by search engines, and is commonly used as a black market of sorts, worthwhile data is usually not simple to come by.

Following close collaboration with KELA, a leading provider of Dark Net threat intelligence, I will examined several methods of machine learning on large data sets mined by KELA's systems. The data includes a wide range of topics such as unstructured hacker discussions, log files from network vulnerability scanners, and other types of information not typically found anywhere else.

The ultimate goal is to apply machine learning to numerous Dark Net sources in order to find trends in attack vectors, and to predict the probability of cyber attacks which use these vectors.

Bio

Sigalit Bechler, passionate and fascinated with science and the understanding of nature and its governing laws.

Sigalit has a diverse academic background; B.Sc's in electrical engineering and in physics, M.Sc in Physics, and two years of Ph.D. research in bioinformatics, after which she decided to join the fascinating world of data science in Israel's hi-tech industry. Sigalit believes in the power of the infusion of ideas and techniques from different fields, and has been working over the past few years as a researcher, data scientist and algorithm developer in several of Israel's market leading hi-tech companies.

Tutorial

NLP Tutorial

Hila Gonen,
PhD Student at Bar-Ilan University in the Field of NLP and Deep Learning

Abstract

Word embeddings have become extremely popular in NLP as they proved useful for a wide range of tasks and are easy to incorporate into neural-based models. Recently, many advances were made in this field, such as introducing multi-sense word embeddings and multilingual word embeddings.

In this tutorial we will review the main algorithms for creating meaningful word embeddings, learn how to use them, and discuss the challenges of evaluation.

Bio

Hila is a PhD student at Bar-Ilan university in the field of NLP and deep learning, under the supervision of Dr. Yoav Goldberg. She is fascinated by languages and is interested in relations between different languages and the way multilingual signals can be used for various tasks. Lately, she has been focusing also on research in the area of code-switching.