Next Innovation in Pharma - CDISC data and Machine Learning

Kevin Lee, Clindata Insight, Moraga, CA

ABSTRACT

The most popular buzz word nowadays in the technology world is "Machine Learning (ML)." Most economists and business experts foresee Machine Learning changing every aspect of our lives in the next 10 years through automating and optimizing processes. This is leading many organizations including drug companies to explore and implement Machine Learning on their own businesses. The presentation will discuss how Machine Learning can lead the next innovation in pharma with CDISC data.

The presentation will start with the introduction of most innovative companies and how they innovate and lead the industry using Machine Learning and data. Then, the presentation will show how pharma should learn from them to innovate using Machine Learning and CDISC data.

The presentation will also introduce the basic concept of machine learning and the importance of data. In the Machine Learning/AI driven process, data is considered as the most important component. 80 to 90 % of works in Machine Learning is preparing the data. Since FDA mandated CDISC standards submission, all the clinical trial data are prepared in CDISC SDTM and ADaM data format. The presentation will show how CDISC data will be the perfect partner of Machine Learning for the next innovation in pharmaceutical industry. Finally, the presentation will discuss how biometric department can prepare the next innovation and lead this data-driven Machine Learning process in pharmaceutical industry.

EXPONENTIAL ORGANIZATIONS

Recently we have witnessed many organizations to grow and to expand exponentially. For example, Uber started their business in 2009 and now in 10 years, its market valuation is about $120 billion. Such organizations are called exponential organizations because they have grown and scaled their business exponentially. The interesting points about the exponential organizations are what they use to grow their business. Rather than using people and physical capacities, the exponential organizations use emerging technologies – data and analytics (mainly Machine Learning). For example, Uber does not own any vehicles, but they use data and analytics to provide access to those who need transportations. And currently, they occupy 70% of hailing business in USA.

HOW EXPONENTIAL ORGANIZATION USE DATA AND ANALYTICS
Exponential organizations use data and analytics differently from the traditional organizations. For exponential organizations, data and analytics are the most important assets. The exponential organizations use data and analytics to make their products and service better, which will attract more customers, obtain more data, build better algorithms, then finally using data and better algorithms, they make better products and service.   This is how "data and analytic cycle" helps organizations to thrive in the era.

More Users

More Data

Better Algorithms (ML/AI)

Better Products
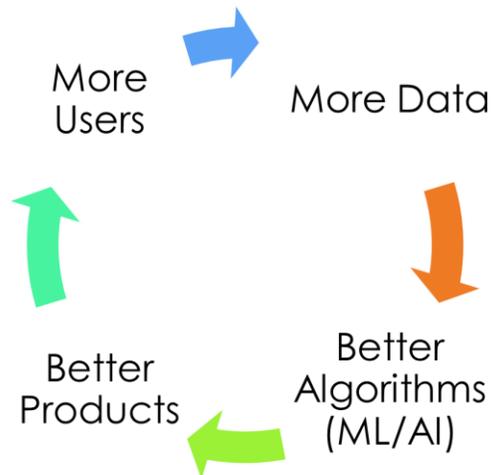
WHY MACHINE LEARNING?
For last couple of year, Machine Learning have gone from futuristic theory to mainstream capabilities. All those innovative companies like exponential organizations such as Google, Amazon, Apple, Microsoft, Netflix and IBM leverage and use Machine Learning to grow their business. And Machine Learning becomes their top priority business assets and tools.

INTRODUCTION OF MACHINE LEARNING
Machine Learning is an application of artificial intelligence (AI) that provides systems or machines the ability to automatically learn and improve from experience without being explicitly programmed whereas AI is human-made or non-natural things that can understand or think.
The key of the Machine Learning definition is "**without being explicitly programmed**".   What it means is that the machines will learn new skills by themselves without programmers adding new rules for new skills.

WHY MACHINE LEARNING IS SO POPULAR NOWADAY
The machine learning is predicted to revolutionize all the industries, especially healthcare industry. First, the Machine Learning can help us to be innovative, solving a lot of complex business problems that we have not been able to solve before and expanding to the area that we were not able to reach.

Secondly, Machine Learning can be also very cost effective since it can automate a lot of mundane and repetitive process. Andrew Ng, the founder of Coursera and previous chief data scientist of Baidu said, **"Pretty much anything that a normal person can do or think less than 1 second, we can now automate with AI"**. Machine Learning along with robotic is expected to automate a lot of human labors. According McKinsey, **as many as 375 million workers (14% of global workforce) may need to switch jobs due to automation by Machine Learning and AI**.

Due to its innovative nature and cost-effective potential, many companies and thought leaders are considering Machine Learning as the next industrial revolution. More and more businesses are implementing Machine Learning to innovate and lead the next industrial revolution.

THE CURRENT MACHINE LEARNING IMPLEMENTATION
Machine Learning is being used more than we realized. The followings are Machine Learning implementation in our daily lives.
- Voice Recognition System – Siri, Alexa, Google Home
- Recommendation – Amazon, Netflix, Spotify
- Customer Service – Online chatting (e.g., Chatbots)
- Cashless store – Amazon GO
- Autonomous Vehicles – Tesla, Google
- Image recognition - CT scans
- Face recognition

MACHINE LEARNING IMPLEMENTATION IN PHARMACEUTICAL INDUSTRIES
Machine Learning is believed to have a huge impact in pharmaceutical industry since 1/3 of data comes from healthcare industry. Its implementation is still in fancy, but more and more pharmaceutical companies go into this area. The followings are some use cases or initial investments and partnerships of pharmaceutical companies on Machine Learning implementation.
- GSK signed 43 million-contract with Exscientia to speed drug discovery. GSK aims to reduce ¼ time (i.e., 5.5 to 1 years) and cost to identify a target for disease intervention to a molecule.
- Surgical Robotics in J&J partners with Google. It will leverage AI/ML to help surgeons by interpreting what they see or predict during surgery.
- Roche works with GNS healthcare to use ML to find novel targets for cancer therapy using cancer patient data.
- Pfizer works with IBM and utilize Watson for drug discovery. Watson has accumulated data from 25 million articles compared to 200 articles a human researcher can read in a year.
- Novartis partners with Watson to develop a cognitive solution using real-time data to gain better insights on the expected outcomes. With Cota Healthcare, Novartis also aim to accelerate clinical development of new therapies for breast cancer.

Many pharmaceutical companies try to utilize AI/ML to automate and speed up the following areas.
- Drug discovery
- Drug candidate selection
- System optimization
- Medical image recognition
- Medical diagnosis
- Optimum site selection or recruitment
- Data anomality detection
- Personized medicine
- Medical coding
- Maybe SDTM, ADaM and TLF developments?

DATA AND MACHINE LEARNING IMPLEMENTATION IN PHARMACEUTICAL COMPANIES.
The question is now whether Pharmaceutical companies can become exponential organizations. The answer is "Yes". Pharmaceutical companies can learn from exponential organizations. We can use data and Machine Learning to grow our business and contribute more to our society.

DATA AND MACHINE LEARNING IMPLEMENTATION IN PHARMACEUTICAL COMPANIES.
One of the most important assets for exponential organizations are data and Pharmaceutical companies have data. For examples, Pharmaceutical companies have clinical trial data. Since FDA mandated clinical trial data to CDISC standards as of Dec 17th, 2017, all pharmaceutical companies have CDISC clinical trial data. The CDISC clinical trial data is perfect for Machine Learning implementation.
- It is clean. Pharmaceutical companies spent a lot of times and resources to clean the data.
- It is unbiased. Clinical studies are prospective studies and many of them are randomized.
- It is blinded. Many pivotal studies are blinded.
- It is structured. CDISC data is structured.

- It has metadata. CDISC data has a metadata, which means that it is much easier for people to understand about data.
- It is standardized. CDISC is standards and any data integration will be a lot easier for Machine Learning implementation.

Just like any other projects, Machine Learning projects also require a lot of data preparation. Usually, data preparation takes about 80 to 90% of all the project. So, CDISC data is perfect data for Machine Learning training since it has been prepared.

## THE REALITY OF CLINICAL TRIAL DATA
However, currently the reality of clinical trial data is not bright.
- Its main purpose is for regulatory submission.
- No or limited analysis after submission
- Usually, people do not know where clinical trial data is
- Limited and sometime no access to clinical trial data for further analysis.
- No central data repository for clinical trial data

Biometric department can play a major role to utilize one of the most important assets (CDISC data) in our organization. We live with CDISC data and know how to use them.

THE FUTURE OF CLINICAL TRIAL DATA USING MACHINE LEARNING
Pharmaceutical companies also can utilize Machine Learning and CDISC data to provide better service and products to healthcare industries.
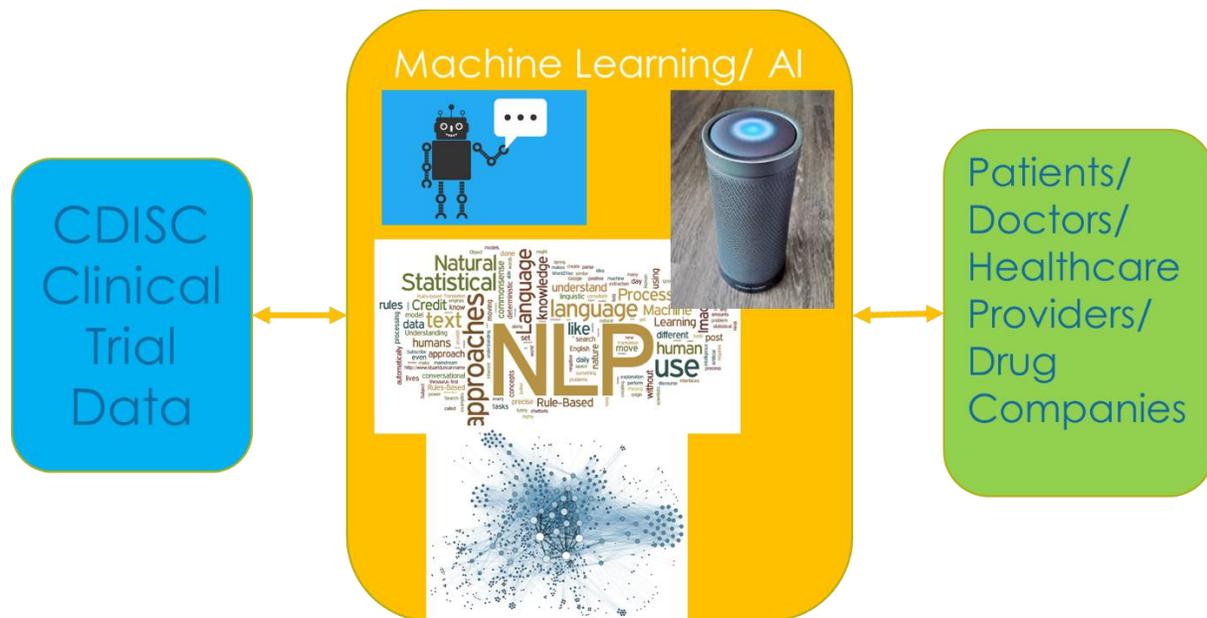
CONCLUSION

Apple is one of the most innovative and respected company in the world. Apple makes products such as iMac, iPhone, and iPad, but we don't really consider Apple as a manufacturing company.   The reason is simple; Apple is not selling products, but Apple is selling experiences. With its user-friendly and AI-driven apps, Apple is providing the devices that people enjoy a great user-experience with. Apple is doing all these with data and Machine Learning.

I believe that the future of pharmaceutical companies will be very bright if they will provide user-experience along with medicine to healthcare community. It will help pharmaceutical companies to get a respect from healthcare industry and to expand their business more than just selling drugs.

REFERENCES

- Exponential Organizations by Malone Ismail & Van Geest

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Director of Data Science
Clindata Insight
klee@clindatainsight.com

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

 ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.