EAS 504 : Application of Data Science - Industry Overview
Assignment 7
Name : Aboli Rawanhate
UB person number - 50374341

Q1. Describe the market sector or sub-space covered in this lecture.

Retail industry is the market sector or sub-space covered in this lecture by Anurag Bharadwaj. The retail sector is primarily concerned with making money by selling items and services to diverse clients via a variety of channels. Data Science is used in many components in the life cycle of retail products. It is one of the industries in which data science is used in almost every aspect of its operations. Data Science is used in many parts of the retail industry like Data and Analytics, Local Daily Deals, Online to Offline Payments, Marketing Platforms, Advertising Technology, Price and Feature Comparison, Product Recommendations, Retail Augmented Reality, etc. It's an industry with a wide range of data challenges that necessitates expertise in natural language processing, deep learning, machine learning, reinforcement learning, augmented reality, computer vision, and other data science methods. It is possible to make efficient judgments in the retail sector using machine learning since there is such a large volume of data gathered and processed. Data science aids in increasing profitability and efficiency, automating processes, lowering expenses, and so on. Forecasting stock models, enticing customers with appealing promotions and rewards, balancing pricing are all ways to draw people to your items. Companies like Paypal, Shopify, Amazon use data science and machine learning to use user shopping history or bills or payment records to derive solutions to business problems.

**Q2. What data science related skills and technologies are commonly used in this sector?**

The retail industry employs a wide range of data science skills and technology. In this sector, data science is used at practically every level.
- To deal with large amounts of data stored in distributed systems. We use Spark.
- Machine learning is used to tackle problems such as product ranking and categorization.
- Demand forecasting techniques are required to tackle the inventory management problem.
- LSTM is used to predict which commodities will be in higher demand and when.
- For visual search and operations such as extracting text from photos and videos and therefore transforming offline material to digital, OCR and Computer Vision are utilized.
- Deep learning is utilized to evaluate data from numerous perspectives, resulting in a 3-D picture and data from an infrared camera of the object being created.

- Warehouse planning uses data science technology to determine the best site for a warehouse, taking into consideration aspects such as demand in the region, proximity to a port, transportation, land availability, warehouse construction costs, and so on.
- In warehouse operations, machine learning and reinforcement learning are utilized to transfer merchandise from one location to another.
- For projecting what goods should be maintained in stock, machine learning is applied. Demand modeling employs STL decomposition.
- Product suggestions in retail shopping sites are made using collaborative filtering tools.
- To give users with a better service experience, machine learning and natural language processing (NLP) are utilized to create automated chatbots.
- Fraud detection for payments etc. Machine learning is used.
- Natural language Processing is used to interpret user query and map to its solution.
- Customers may experience the look of a product in their environment using Augmented Reality technology, allowing them to make a better decision on which product to buy.
- Power BI and Tableau are data visualization tools that enable you to see how elements like logistics, retail payments, and ads evolve over time.

Q3. How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

We will see data and computing related methods used in the retail industry by observing the Life Cycle of a T-shirt.
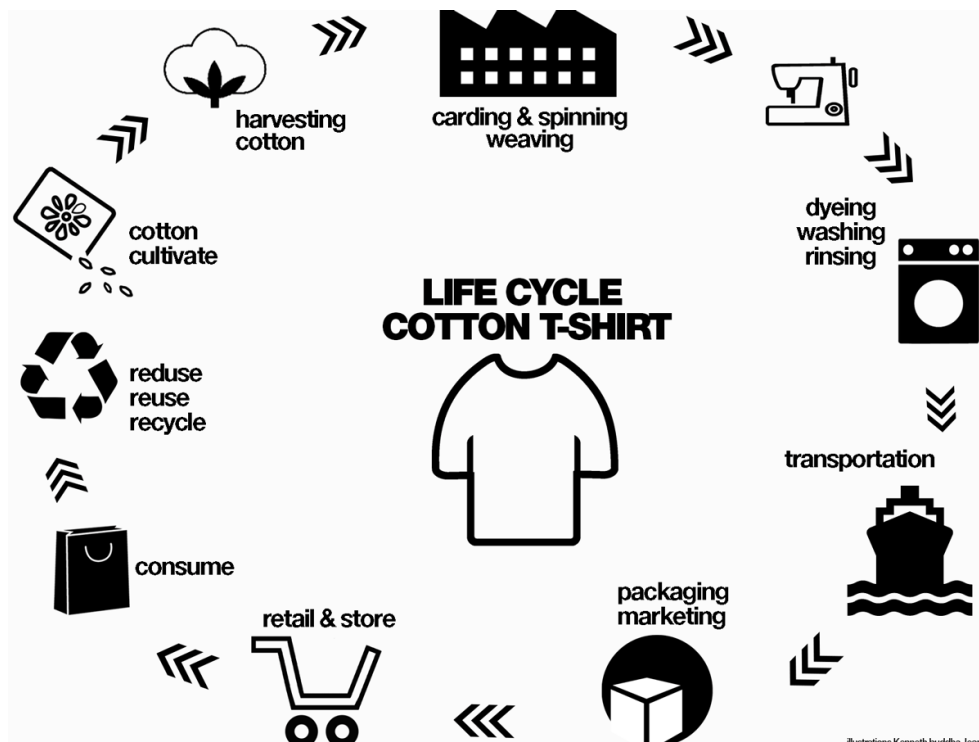
There are five phases in the Tshirt retail life cycle:

1. The material phase of the life cycle involves farming, irrigating, fertilizing, harvesting and ginning. Cotton is irrigated, fertilized, and cultivated on fields where it is grown and harvested. Then it is transported to a warehouse where it will be utilized in the fabrication of T-shirts. Poor pricing, minimal retailer margins, and low customer awareness characterize this stage.

2. The production phase begins once the cotton is grown and collected. Spinning, knitting, wet process, bleaching, dyeing, confection, cutting, and sewing are all techniques used in the textile industry. Commercial dyes and bleaches are dangerous pollutants that can contaminate groundwater, in addition to the massive amounts of water and energy used in all of these activities. This phase will have little infrastructure at first, but as demand grows, it will expand.

3. The t-shirt must be transported after it has been created. You'll probably discover a range of nations of origin on your clothes tags if you peek in your closet. According to studies, freight vehicles alone emit 1.15 pounds of carbon dioxide every mile into the atmosphere. Much more is used by planes, ships, and railways. This phase most of the infrastructure is invested in transportation and making the product known to customers.

4. The next phase is retail, which may seem like the least harmful phase but… think again. The biggest environmental drain your clothing causes comes from the repeated washing and drying we all do to keep our clothes clean. The merchant competes with other retailers, opens new locations, and advertises his items in an attempt to extend his network. Larger pricing, higher retailer margins, and improved consumer awareness characterize this phase, which boosts the retailer's reputation.

5. The final stage for the t-shirt is disposal. Reusable items are recycled and re-introduced into the market. The decline in investment and infrastructure characterizes this stage.

Q4. What are the data science related challenges one might encounter in this domain?
Data science related challenges one might encounter in this domain are:

● Customizing such bargains depending on their interests, age, geographic area, things purchased more frequently in that region, and other factors is a difficult task for data scientists to solve.

● False positives may merely harm yield, but false negatives may damage product quality.

● Adhering to data protection rules is getting more difficult as a result of numerous data breaches caused by hackers.

● Retailers must keep up with the competition or risk losing business and having to close their doors.

● Because retail is complicated, diversified, and comprehensive, even basic problems demand new AI/ML solutions.

- It's difficult to make good forecasts because low-quality data leads to erroneous conclusions.
- There are several sources of sales volumes, client information, profit margins, stock levels, and advertisement marketing, resulting in a wide range of outcomes.
- Recommending the correct product to the user is the biggest challenge since there are so many variables to examine before suggesting a product to a buyer.
- Searching something on the site and getting desired results is one of the most challenging issues to overcome because if search fails, the whole user experience is jeopardized.
- Programming robots accurately to maintain supplies in a warehouse is vital and may be difficult. Because product demand fluctuates depending on the time of year, seasonality and data trends must be considered when building any data science system for forecasting, which can be difficult.

Q5. What do you find interesting about the nature of data science opportunities in this domain?

Data science skills and technology are widely employed in every element of the retail business, including production, warehousing, inventory management, price optimization, search, discovery, shopping experience, trust and safety, and customer service.

- In order to keep on top of demand, seasoning or trend checking is done. If there is a shift, the trend is altered, which is known as de-seasoning. There is no need to process anything else.
- To calculate the best pricing for a product, NLP and machine learning are utilized. Whether it comes to selecting when to use a clearance sale in order to maximize earnings, selling through optimization can assist.
- Advanced machine learning models, such as LSTM, are employed when several inputs and multiple outputs are required.
- In robotic automation, experts in technologies such as deep learning, computer vision, and reinforcement are employed. In inventory management forecasting, machine learning is commonly employed.
- Depending on the task at hand, many types of forecasting are used, such as naive forecasting, statistical forecasting, demand planning, demand modeling, and machine learning.
- Statistical performance measures are calculated to stay on top of the trend and assess current outcomes.
- The requirement is raised or lowered depending on the seasoning. The results are compared to logs that have been noticed before.

- Customers can upload images and search results relevant to the image using computer vision and image processing.
- Providing dynamic pricing and personalized ads to attract customers.
- Recommending products favorable to customers resulting in increased revenue.

Q6. Describe some of the data science problems relevant to Manufacturing and Warehousing in the Retail Product Lifecycle.

Manufacturing:
- Automating production of products using robots. This process has to be integrated with all other processes. Interconnecting pins through golden wire in semiconductors is done by robotic tools.
- Constructing chips if it takes 10 secs then assessing its quality will take 30 secs. It can be minimized using machine learning. To detect flaws, ML is used.
- But after getting an error, the yield number is reduced. So if we have a lot of false positives, the yield number is reduced. False negative is it was a faulty chip but detected as non-faulty. These are the data science problems we face in Manufacturing. We can't have false negatives.
- Microscope images of the chip are also read using Computer vision and faults detected in chip
- Self driving cars - We need a physical map, it is static. Then we need a dynamic map as well. Are we having trees, animals, buildings around? Any obstacle. So combining both the maps we have a reality created around us. Robotic automation, AR is used in this. Motion and planning: Now I have these things, what should I do? Shall I change speed? Shall I apply for a break, etc. Reinforcement Learning is used to make decisions.

Warehousing:
Lots of warehouse problems are solved using machine learning.
- Warehouse planning - It involves data analysis that it should be close to railroads, accessible to the stores, etc. So, data analysis helps us to get insights on where the warehouse should be located.
- Warehouse operation - Transport objects from assembly point to warehouses. Millions of requests are made on festival days, and robots come into action. As soon as demand comes, they fetch the product from the warehouse, put on inventory, put the label on it, and place it into a shipping container. There is always a truck available to get these items and thus the complete shipping process is handled by robots without manual intervention. Thus robotic automation is used here. This avoids accidents.

- Inventory Management - What to store in a warehouse, as per demand in that region, inventory management is done using machine learning.
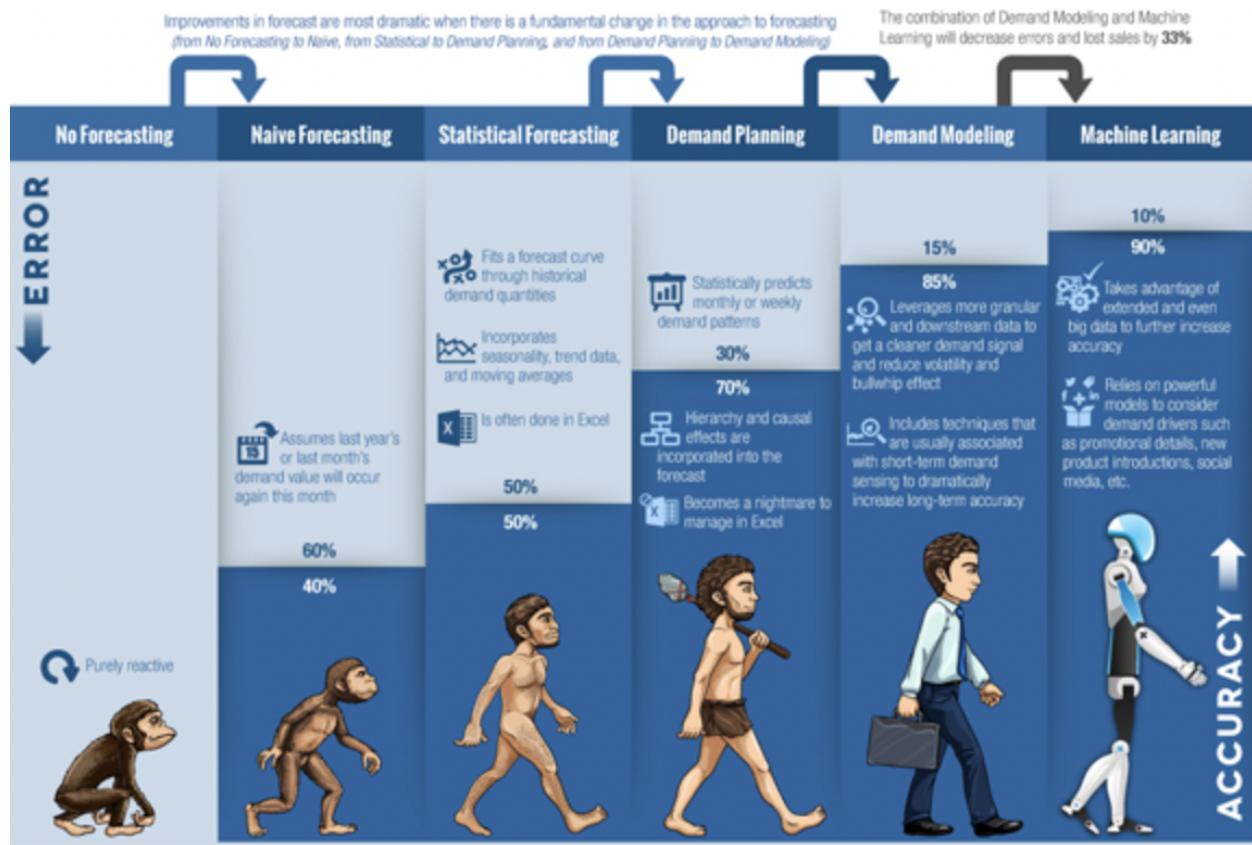
Q7. Describe some of the data science problems and techniques that would be useful in Inventory Management and Pricing Optimization

Data science problems and techniques that would be useful in Inventory Management and Pricing Optimization are as follows:

Inventory management
- Biggest question in inventory management is what items should be kept in inventory? As people have many choices on clothes and phones, it's not possible to keep them all. So, which line of clothes should be kept is one of the problems solved using data science.
- Should it be sent only online or offline? - Is it possible to keep items offline? Will it cost more ? These questions are answered by machine learning.
- What should be kept in stores, what should be kept in warehouses. Distance for shipping is less from store than warehouse then it's directly shipped from store. Companies like Walmart used machine learning to work on such problems.
- Based on demand, what should be the price of a product in a particular region?

Forecasting has evolved from naive forecasting then to statistical forecasting where statistics were used to predict the demand and prices, then demand planning came into picture. To make it better, demand modeling was used and finally machine learning gave best results and hence used in forecasting.

Improvements in forecast are most dramatic when there is a fundamental change in the approach to forecasting *(from No Forecasting to Naive, from Statistical to Demand Planning, and from Demand Planning to Demand Modeling)*

The combination of Demand Modeling and Machine Learning will decrease errors and lost sales by **33%**

ERROR

No Forecasting	Naive Forecasting	Statistical Forecasting	Demand Planning	Demand Modeling	Machine Learning
Purely reactive	Assumes last year's or last month's demand value will occur again this month	Fits a forecast curve through historical demand quantities; Incorporates seasonality, trend data, and moving averages; Is often done in Excel	Statistically predicts monthly or weekly demand patterns; Hierarchy and causal effects are incorporated into the forecast; Becomes a nightmare to manage in Excel	Leverages more granular and downstream data to get a cleaner demand signal and reduce volatility and bullwhip effect; Includes techniques that are usually associated with short-term demand sensing to dramatically increase long-term accuracy	Takes advantage of extended and even big data to further increase accuracy; Relies on powerful models to consider demand drivers such as promotional details, new product introductions, social media, etc.
	60% / 40%	50% / 50%	30% / 70%	15% / 85%	10% / 90%

ACCURACY

Pricing and Optimization:

- Prices can be kept different online and offline. Items are expensive to keep in store then prices will be different.
- To keep prices consistent between manufacturers and customers, deep learning is used
- The active learning algorithm aids in the design of appropriate stock and inventory strategies by continually altering and developing parameters and values
- Chatbots can be used to gain a deeper understanding of each individual consumer
- Machine learning models are used to calculate statistical performance measures in order to stay up with the trend and analyze current outcomes.

Q8. MCQs

Q1. B

Q2. C

Q3. D

Q4. B

Q5. C