An Overview of Machine Learning

By: Aditi Chaudhari

I. What is Machine Learning?

a. Definition of Machine Learning

Machine learning is a field in computer science that uses algorithms to analyze data, recognize patterns, and then make accurate predictions on data that the algorithm has not seen yet.

b. The importance of data, pattern recognition, and accuracy

Data is incredibly important to machine learning. It can be stored in many different formats and is either structured or unstructured. Structured data is data that is highly organized and stored in a predefined format. Examples of structured data could be first and last names, age, and sex. On the other hand, unstructured data is data that does not have a predefined structure and can come in many different data formats. Examples of unstructured data could be text files, video files, and audio files. Using data, a computer can be trained to recognize patterns. Raw data is carefully converted into organized data and then decisions are made to determine which patterns are important and what can be learned from the data. This pattern recognition allows us to make predictions for future data. It is important for these predictions to be accurate, though, because if they are not accurate, they are simply random guesses. There are techniques in place to determine how accurate an algorithm is and how well it has performed.

c. The relationship between artificial intelligence and machine learning

Artificial intelligence and machine learning are closely related, yet not the same. Artificial intelligence refers to the general ability for computers to mimic human thought and behavior. Machine learning is a subfield of artificial intelligence that recognizes patterns in data and is able to make predictions based on these patterns. Essentially, machine learning is one of the subfields that paves the way for artificial intelligence.

d. Definition of observation, feature, qualitative data, and quantitative data

A lot of data used in machine learning is stored in tables. Each row in the table is called an observation, whereas each column in the table is called a feature. It is also important to distinguish between qualitative data and quantitative data. Qualitative data is descriptive in nature and cannot be easily counted, measured, or quantified using numbers. If we were to look at a student, an example of qualitative data could be their class (whether they are a freshman, sophomore, junior, or senior). On the other hand, quantitative data is numeric data. A student's GPA, the hours they're taking, and their SAT score are all examples of quantitative data. It is important to distinguish between observations and features and also qualitative and quantitative data because different features can be used to predict a target variable. By knowing which data is an observation and which data is a feature, a developer will be better equipped to write algorithms to predict a target variable. Similarly, knowing the difference

between qualitative and quantitative data allows the developer to know which types of algorithms to use.

II. Examples of Machine Learning

An example of a modern machine learning application is Netflix and how it provides recommendations on what you should watch based on what you've already watched and enjoyed. The algorithms that Netflix recommendations use cannot be written with traditional programming because it is not possible for a software developer to encode all of the rules necessary to provide a recommendation. For instance, if a user watched and enjoyed a sitcom named *The Office*, then the software developer could attempt to encode other sitcoms to recommend, such as *Friends, Parks and Recreation, Community,* and more. However, the software developer would have to do encode rules for every single movie and TV show on Netflix, which would be incredibly challenging. The fact that the movies and TV shows on Netflix change every month would make it impossible. Therefore, it'd be better to use a machine learning algorithm to accurately predict what Netflix's users want to watch.

Another example of a modern machine learning application is Google Maps. Whenever you leave your home, you probably check Google Maps (or another navigation application) in order to find the fastest route to your destination. Google Maps cannot provide the fastest route to a user's destination using traditional programming because the scale of the problem is too large. There is simply too much data on all of the routes a user could take to get to their destination based on traffic accidents, slowdowns, and more. A human wouldn't be able to take in all of this data and find the fastest route possible, which is why its better for Google Maps to use a machine learning algorithm that can take in all the data, recognize patterns, and predict the fastest possible route to the user.

III. My Personal Interest in Machine Learning

I am fascinated in how applications that I use everyday extract data from me and use it to predict who I am and what I want. For instance, Amazon has picked up on the fact that I am a student based on what I've bought in the past, and now recommends different desks and stationery to me. Spotify has recognized that I like listening to hip hop, R&B, and indie music and consequently created a playlist for each genre. TikTok has picked up on my cultural identity and recommends videos from other creators of the same cultural background. I can list so many other examples, as well.

I simultaneously feel excited and scared by what machine learning algorithms can learn about us. My excitement is what has led me to learn more about machine learning, so that I can understand how these big companies extract data from us and use it to predict who we are and what we want. My fear, on the other hand, has led me to want to learn more on the ethics of data collection and machine learning. I hope that my excitement will guide me to be a better software engineer, incorporate machine learning algorithms into any personal projects I pursue, and encourage me to work at a company that is a pioneer in the fields of artificial intelligence and machine learning. I hope that my fear keeps me grounded and serves as an important reminder to consider how the products I build affect other people.