Practical - 1

2CS702 – Big Data Analytics

Harshit Gajipara
19BCE059

Aim:

Study and explore various applications of big data in different domains. Choose one of it and study in detail, also write down the report on different types of digital data generated in selected application.
For e.g.:
- Big Data in Retail
- Big Data in Healthcare
- Big Data in Education
- Big Data in E-commerce
- Big Data in Media and Entertainment
- Big Data in Finance
- Big Data in Travel Industry
- Big Data in Telecom
- Big Data in Automobile

Big Data:

Big Data meaning a data that is huge in size. Big data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.

Big Data, a term that has lately gained popularity, is now understood to refer to a sizeable volume of data that cannot be stored or handled using standard data storage or processing tools. Due to the enormous volume of data generated by both human and machine activity, the data are so complicated and vast that neither people nor a relational database can analyse them. However, when properly analysed using contemporary techniques, these vast amounts of data offer firms significant insights that aid them in improving their operations by taking wise decisions.

Types of Big Data

Following are the types of Big Data:

1. Structured
2. Unstructured
3. Semi-structured

Structured Big Data

Structured data refers to any data that can be accessed, processed, and stored in a fixed format. Over time, computer science expertise has had more success creating methods for handling this type of data (when the format is fully understood in advance) and also extracting value from it. Today, we are anticipating problems as the size of this data increases significantly; average sizes are now in multiple zettabytes.

Ex. Data stored in RDBMS

Unstructured Big Data

Unstructured data is any data whose structure or pattern is unknown. Unstructured data is enormous in quantity and presents a number of processing obstacles that must be overcome in order to extract value from it. Unstructured data is frequently found in heterogeneous data sources that combine simple text files with photos, videos, and other types of data. Organizations nowadays have a lot of data at their disposal, but since this data is in its raw or unstructured form, they are unable to value-add from it.

Ex. Data generated from Facebook, webpages returned from google search etc.

Semi-structured Big Data

Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data is a data represented in an XML file.

Ex. Personal data stored in an XML file

Characteristics of Big Data

Big data can be described by the following characteristics:

- Volume – size of data which is obviously huge
- Variety – types of data being generated
- Velocity – speed in which data is being generated

These 3 characteristics define if the given problem is a big data problem or not. If the data is not having all these characteristics then the data given is not big data.

Big Data in Agriculture

India is one of the most extensive agricultural countries in the world. There are a total of 29 states in India. All the states have their own version of climate and terrain conditions. Some states have a portion of coastal areas, whereas some consist of desert or mountain areas. Due to these geographic variations, many types of crops are being grown in different states of India. Agriculture is the backbone of the Indian economy and contributes up to 23% to Indian Gross Domestic Product (GDP). But most of the agricultural activities depend only on rain and the rainfall conditions are different in every region. So farmers are heavily dependent on the rainy season.

Farming depends on various factors such as weather, availability of water, quality of land, seeds, and pesticides. After the seeds have been planted, a scientific way of farming needs to be incorporated in order to intensify the yield. After receiving the appropriate yield, a proper market is required to get the best selling price for the crops. The scientific and technical way of farming may lead to quality yields at the end.

Applications of Big Data in agriculture

Big data applications have a broad range, and we've just just scratched the surface of their potential. A true game changer in agricultural techniques could be the capacity to track tangible objects, gather real-time data, and foresee possible outcomes. Let's look at a few instances where using big data might be beneficial.

1. Providing food for an expanding population

Even governments are working together to find a solution to one of the major problems. The yield of currently used farmlands can be increased as a means of achieving this.

Granular information on rainfall patterns, water cycles, fertiliser needs, and other topics is made available to farmers through big data. They can use this information to make wise choices about when to harvest and what crops to grow for maximum profit. Making the proper choices ultimately increases crop output.

2. Making ethical use of pesticides

The use of pesticides has generated controversy because of its adverse effects on the environment. With the use of big data, farmers can better manage this by being advised on when, where, and how much pesticides to use.

Farmers can follow rules set by the government and limit the use of chemicals in food production by regularly monitoring it. Additionally, because crops aren't harmed by weeds and insects, this increases profitability.

3. Improving farm machinery

Businesses like John Deere have implemented big data apps and integrated sensors into their farming machinery in order to better manage their fleet. This kind of monitoring, which notifies users of tractor availability, service due dates, and fuel refill warnings, can be a lifesaver for large farms. In essence, this maximises usage and guarantees the durability of farm machinery.

4. Addressing issues with the supply chain

According to McKinsey, every year, a third of the food produced for human use is lost or squandered. An unfortunate reality given the industry's attempts to close the supply-demand mismatch. Food supply cycles from the producer to the market must be shortened to address this. Using big data to improve supply chains

Big data in a general sense is defined using 3 V's - Volume, Velocity, and Variety. In our case, there is a large amount of data (around Petabytes or more). The data captured for weather and for crop's humidity and moisture level are continuous in particular intervals of time. So we can say that the velocity of data generation is high.

Types of Data in Agriculture

Agriculture data contains data relating to various elements such as weather, geography, crop diseases, pesticides, and market price for different crops, etc.

- Weather data
- Geographic data
- Diseases data

- Data from sensors
- Selling price data
- Data of pesticides

Many government or company websites are generating weather data for every hour or every minute. One can use weather data to decide the farming of crops according to their geographic region, prediction of rain, hotness, and coldness.

Geographic data can be used to decide the quality and type of land for crops. During farming, there may be some diseases in the crop. These diseases can turn into less yield in the end. So we can set some sensor cameras to capture the images of those diseased crops.

Some crops need perfect humidity, moisture, etc. to get more yield. For this, we can use some sensors that take data in particular intervals of time. Using that data, we can generate a report on the humidity and moisture.

For selling the crops, we need to decide the best market price. So we have many options as there are multiple markets with different price for different crops. On that data, we can create different charts like crop-wise price, market-wise price, daily market report, price trend, etc. to analyse the market prices and to predict the future trend of market prices.

Data	Type
Photos of crop diseases	Unstructured
Market Price of Crops	Structured
Geographic data on fields	Semi structured
Data from sensors	Semi structured
Pesticide data	Unstructured
Weather data	Unstructured