DATA ANALYTICS IN IOT

Kamil Khan

Ezhilarasi

Chiranjeevi

Dhavakumar

Veena

Table of Contents

- IOT Analytics and Future of IOT Data Analytics - Kamil

- Future of IOT Analytics, Job Prospects and Dark Data - Ezhilarasi

- Softwares For Data Analytics - Open Source and Commercial - Veena

- Features and Types of Analytics - Chiranjeevi

- Applications, Use cases, Challenges - Dhava Kumar

IOT Analytics

- By 2025, forecasts suggest that there will be more than 75 billion Internet of Things (IoT) connected devices in use. This would be a nearly threefold increase from the IoT installed base in 2019.

- IoT analytics is the analytics platform that can assess the data collected from IoT devices.

- IoT data is similar to big data but there are differences not just in terms of size, but also because of the diversity of sources. The heterogeneous data sources make data integration an incredibly complex process

- Many organisations do struggle with incorporating and handling IoT because they simply lack the right systems that can handle the volume, velocity and variety of IoT data.

- IoT analytics is incredibly versatile and can be used for any purpose.

- Both for-profit and non-profit organisations will benefit from this version of the data analytics platform.

Advantages of IOT Analytics

❖ Rigours of analysing IoT data

❖ Giving organisations insight into operational efficiency

❖ A better understanding of the market

❖ Competitive Advantage over other competitors

Future of IOT Analytics

❖ Chief Data Officers (CDO) - Many companies are incorporating this Role.

❖ According to Forrester, emergence of CDOs as a new and urgent position that needs to be filled.

❖ Global IoT market will grow at a rate of 28.5%.

Dark Data

❖ 90% of unstructured data is never analysed. - Dark Data

❖ Dark data is a subset of big data, which constitutes the biggest portion of the total volume of big data collected by organisations in a year.

❖ Dark data is not processed due to various reasons but that it might have lot of important business value.

❖ Massive strides in trying to collate and decipher the unstructured data available.

Data Analytics Open Source

❖ R - R is the leading analytics tool in the industry and widely used for statistics and data modeling. It can easily manipulate your data and present in different ways.

❖ Tableau Public - Tableau Public is a free software that connects any data source be it corporate Data Warehouse, Microsoft Excel or web-based data, and creates data visualizations, maps, dashboards etc. with real-time updates presenting on web.

❖ Python is an object-oriented scripting language which is easy to read, write, maintain and is a free open source tool. It handles text data very well

❖ Apache Spark - Apache Spark is a fast large-scale data processing engine and executes applications in Hadoop clusters 100 times faster in memory and 10 times faster on disk. Spark is built on data science and its concept makes data science effortless. Spark is also popular for data pipelines and machine learning models development.

❖ KNIME - KNIME Developed in January 2004 by a team of software engineers at University of Konstanz. KNIME is leading open source, reporting, and integrated analytics tools that allow you to analyze and model the data through visual programming, it integrates various components for data mining and machine learning via its modular data-pipelining concept.

Data Analytics – Commercial Solutions

- SAS - Sas is a programming environment and language for data manipulation and a leader in analytics, developed by the SAS Institute in 1966 and further developed in 1980's and 1990's. SAS is easily accessible, managable and can analyze data from any sources.

- Splunk is a tool that analyzes and search the machine-generated data. Splunk pulls all text-based log data and provides a simple way to search through it, a user can pull in all kind of data, and perform all sort of interesting statistical analysis on it, and present it in different formats.

- Excel - Excel is a basic, popular and widely used analytical tool almost in all industries. Whether you are an expert in Sas, R or Tableau, you will still need to use Excel. Excel becomes important when there is a requirement of analytics on the client's internal data. It analyzes the complex task that summarizes the data with a preview of pivot tables that helps in filtering the data as per client requirement. Excel has the advance business analytics option which helps in modelling capabilities which have prebuilt options like automatic relationship detection, a creation of DAX measures and time grouping.

- Rapid Miner - RapidMiner is a powerful integrated data science platform developed by the same company that performs predictive analysis and other advanced analytics like data mining, text analytics, machine learning and visual analytics without any programming.

Analytics Tools – Features

❖ Volume: There are huge clusters of data sets that IoT applications make use of. The business organizations need to manage these large volumes of data and need to analyze the same for extracting relevant patterns. These datasets along with real-time data can be analyzed easily and efficiently with data analytics software.

❖ Structure: IoT applications involve data sets that may have a varied structure as unstructured, semi-structured and structured data sets. There may also be a significant difference in the data formats and types. Data analytics will allow the business executive to analyze all of these varying sets of data using automated tools and software.

Types of Analytics

❖ Streaming Analytics: This form of data analytics is also referred as event stream processing and it analyzes huge in-motion data sets. Real-time data streams are analyzed in this process to detect urgent situations and immediate actions. IoT applications based on financial transactions, air fleet tracking, traffic analysis etc. can benefit from this method.

❖ Spatial Analytics: This is the data analytics method that is used to analyze geographic patterns to determine the spatial relationship between the physical objects. Location-based IoT applications, such as smart parking applications can benefit from this form of data analytics.

❖ Time Series Analytics: As the name suggests, this form of data analytics is based upon the time-based data which is analyzed to reveal associated trends and patterns. IoT applications, such as weather forecasting applications and health monitoring systems can benefit from this form of data analytics method.

❖ Prescriptive Analysis: This form of data analytics is the combination of descriptive and predictive analysis. It is applied to understand the best steps of action that can be taken in a particular situation. Commercial IoT applications can make use of this form of data analytics to gain better conclusions.

Applications of Data Mining in IOT

- Classification - Medical Imaging, Speech Recognition, NLP, e-governance

- Clustering - Health care, Medical Imaging, Human Generics, Market Analysis, Bio Informatics, Social Network Analysis, e-governance

- Association Rules - Health Care, Market Analysis, Bio Informatics, e-governance

- Prediction - Disaster Management, Market Analysis, Social Network Analysis

- Time Series - Disaster Management, Medical Imaging, Speech Recognition, Social Network Analysis, e-governance

Use cases

- ❖ Smart Metering - Predict Electricity consumption - Sensors - Text - Hadoop

- ❖ Smart Transportation - Improve Traffic System - Sensors, cameras - Text, Video, Audio - Hadoop, Spark,Hive

- ❖ Smart Supply Chains - control external environment - Sensors, Mobile Devices - Text, Image - Hadoop

- ❖ Smart Agriculture - Obtain Moisture Level, trunk diameters of plant, forecast weather, microclimate conditions, humidity level - Sensors - Text, Image - Hadoop

- ❖ Smart Grid - Improves reliability, safety, efficiency - Sensors - Text - Hadoop

Challenges

- ❖ Privacy - Privacy issues arise when a system is compromised to inferior restore personal information using big data analytics tools,although data are generated from anonymous users. With the proliferation of big data analytics technologies used in bigIoT data, the privacy issue has become a core problem in the data mining domain. Consequently, most people are reluctant to rely on these systems, which do not provide solid service level agreement (SLA) conditions regarding user personal information theft or misuse.

- ❖ Data Mining - The size and heterogeneity of data impose new data mining requirements, and diversity in data sources also poses a challenge

- ❖ Visualisation - conducting data Visualisation is difficult because of the large size and high dimension of big data

- ❖ Integration - Integrating unstructured data like video, audio, image is still a big challenge.