**POSTLAB 10**

1)Differentiate - Datamining and ML/ BDA

Ans: -

|  |  |  |
| --- | --- | --- |
| basic for comparison | Data mining | Machine learning |
| Meaning | Extracting knowledge from a large amount of data | Introduce new algorithm from data as well as past experience |
| History | Introduce in 1930, initially referred as knowledge discovery in databases | introduce in near 1950, the first program was Samuel’s checker-playing program |
| Responsibility | Data mining is used to get the rules from the existing data. | Machine learning teaches the computer to learn and understand the given rules. |
| Origin | Traditional databases with unstructured data | Existing data as well as algorithms. |
| Implementation | We can develop our own models where we can use data mining techniques for | We can use machine learning algorithm in the decision tree, neural networks and some other area of artificial intelligence. |
| Nature | Involves human interference more towards manual. | Automated, once design self-implemented, no human effort |
| Application | used in cluster analysis | used in web search, spam filter, credit scoring, fraud detection, computer design |
| Abstraction | Data mining abstract from the data warehouse | Machine learning reads machine |
| Techniques involve | Data mining is more of a research using methods like machine learning | Self-learned and trains system to do the intelligent task. |
| Scope | Applied in the limited area | Can be used in a vast area. |

2) Need of deep learning

Ans: -

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It’s achieving results that were not possible before. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labelled data and neural network architectures that contain many layers.

In a word, accuracy. Deep learning achieves recognition accuracy at higher levels than ever before. This helps consumer electronics meet user expectations, and it is crucial for safety-critical applications like driverless cars. Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images.

While deep learning was first theorized in the 1980s, there are two main reasons it has only recently become useful:

1. Deep learning requires large amounts of **labelled data**. For example, driverless car development requires millions of images and thousands of hours of video.
2. Deep learning requires substantial **computing power**. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from weeks to hours or less.

Examples of Deep Learning at Work

Deep learning applications are used in industries from automated driving to medical devices.

Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

Aerospace and Defense: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.

Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

3) Different algorithms and applications

Ans: -

# **Linear Regression**

Linear Regression is widely used for applications such as sales forecasting, risk assessment analysis in health insurance companies and requires minimal tuning.

**Logistic regression**

Logistic regression is used in applications such as-

1. To Identifying risk factors for diseases and planning preventive measures

2. Classifying words as nouns, pronouns, and verbs

3. Weather forecasting applications for predicting rainfall and weather conditions

4. In voting applications to find out whether voters will vote for a particular candidate or not

A good example of logistic regression is when credit card companies develop models which decide whether a customer will default on their loan EMIs or not.

# **K-Nearest Neighbours Algorithm**

The best way to advance our understanding of these algorithms is to try our hand in image classification, stock analysis, and similar beginner data science projects.

# **Support Vector Machine (SVM) Algorithm**

Support Vector Machine Learning Algorithm is used in business applications such as comparing the relative performance of stocks over a period of time. These comparisons are later used to make wiser investment choices.

# **Decision Tree Algorithm**

Applications of this Decision Tree Machine Learning Algorithm range from data exploration, pattern recognition, option pricing in finances and identifying disease and risk trends.

# **Random Forest Algorithm**

The random forest algorithm is used in industrial applications such as finding out whether a loan applicant is low-risk or high-risk, predicting the failure of mechanical parts in automobile engines and predicting social media share scores and performance scores.

# **Naive Bayes Classifier Algorithm**

If we ‘re planning to automatically classify web pages, forum posts, blog snippets, and tweets without manually going through them, then the Naive Bayes Classifier Algorithm will make our life easier. This classifies words based on the popular Bayes Theorem of probability and is used in applications related to disease prediction, document classification, spam filters, and sentiment analysis projects.

We can use the Naive Bayes Classifier Algorithm for ranking pages, indexing relevancy scores and classifying data categorically.

# **Principal Component Analysis (PCA) Algorithm**

PCA algorithm is used in applications such as gene expression analysis, stock market predictions and in pattern classification tasks that ignore class labels.

# **K-Means Clustering Algorithm**

K-Means Clustering Algorithm is frequently used in applications such as grouping images into different categories, detecting different activity types in motion sensors and for monitoring whether tracked data points changes between different groups over time. There are business use cases of this algorithm as well such as segmenting data by purchase history, classifying persons based on different interests, grouping inventories by manufacturing and sales metrics, etc.