Government College of Engineering (GCOEJ), Jalgaon

(An Autonomous Institute of Government of Maharashtra)

DEPARTMENT OF COMPUTER ENGINEERING

INDUSTRIAL LECTURE REPORT ON

MASTERING MACHINE LEARNING

(Academic Year 2023-24)

Submitted by:

Bhargav Shamuvel Gurav (2041009)

Government College of Engineering (GCOEJ), Jalgaon

(An Autonomous Institute of Govt. of Maharashtra)

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the ***Industrial Lecture*** report**, "Mastering Machine Learning"**, which is being submitted here with for the award of ***LY Computer Engineering (7th Semester)*** is the result of the work completed by ***Bhargav Gurav (2041009)*** under my supervision and guidance within offline mode of classes of the institute, in the academic year 2023-24.

.

Head of Department

Dr. D. V. Chaudhari

INDEX

CONTENTS

ABSTRACT

The expedition to master machine learning is an immersive odyssey through the intricacies of data, algorithms, and their applications in crafting predictive models and intelligent systems. This report encapsulates the multifaceted voyage, tracing the trajectory from foundational principles to sophisticated implementations, elucidating the pivotal milestones, challenges, and strategies intrinsic to this transformative journey.

Beginning with a foundational comprehension, this report navigates the foundational concepts, elucidating the fundamental algorithms, and delineating the landscape of supervised, unsupervised, and reinforcement learning paradigms. It explores the symbiotic relationship between theory and practice, elucidating the diverse learning pathways—be it through structured academic courses or the autodidactic pursuit fostered by online resources, interactive platforms, and immersive projects.

Moreover, the report sheds light on the pragmatic aspect of machine learning, illuminating its practical applications across diverse domains—be it healthcare's predictive diagnostics, finance's algorithmic trading, or marketing's customer segmentation. Through illuminating case studies, it elucidates successful implementations, showcasing the transformative power of machine learning in solving real-world challenges.

In tandem, it unravels the tapestry of tools and technologies, spotlighting the significance of frameworks like TensorFlow, PyTorch, and scikit-learn, while emphasizing the indispensability of programming prowess in languages such as Python and R.

In essence, this report encapsulates the holistic expedition towards mastering machine learning, offering a comprehensive roadmap navigating through its myriad facets, empowering and inspiring individuals in their quest for expertise in this transformative domain.

INTRODUCTION

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and

algorithms to imitate the way that humans learn, gradually improving its accuracy.

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix's recommendation engine and self-driving cars.
Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

Machine learning algorithms are typically created using frameworks that accelerate solution development, such as TensorFlow and PyTorch.

Now talking about the purpose of this report then this report is made to describe or to depict the learning path how the tech enthusiastic nerds can land into the field of this vast machine learning world. There are a lot of opportunities for the geeks who are really interested in developing something new, something unique that can help the world to automate, predict and thus making a better perspective for a person to view the world. There can be a lot of ways they can innovate like by making "IRON MAN's goggles" which can do a lot of interesting stuffs like scanning, getting x rays filters, or they can build future predicting machines means there are a lot of possibilities and many more stuff to be created using machine learning and AI.

CONCEPTS

Classification of Machine Learning
Machine learning implementations are classified into four major categories, depending on the nature of the learning "signal" or "response" available to a learning system which are as follows:

A. Supervised learning:
Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. The given data is labeled. Both classification and regression problems are supervised learning problems. Example — Consider the following data regarding patients entering a clinic . The data consists of the gender and age of the patients and each patient is labeled as "healthy" or "sick".

Gender	Age	Label
M	48	sick
M	67	sick
F	53	healthy
M	49	sick
F	32	healthy
M	34	healthy

B. Unsupervised learning:

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. In unsupervised learning algorithms, classification or categorization is not included in the observations. Example: Consider the following data regarding patients entering a clinic. The data consists of the gender and age of the patients.

Gender	Age
M	48
M	67
F	53
M	49

As a kind of learning, it resembles the methods humans use to figure out that certain objects or events are from the same class, such as by observing the degree of similarity between objects. Some recommendation systems that you find on the web in the form of marketing automation are based on this type of learning.

.

C. Reinforcement learning:

Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards.

A learner is not told what actions to take as in most forms of machine learning but instead must discover which actions yield the most reward by trying them. For example — Consider teaching a dog a new trick: we cannot tell him what to do, what not to do, but we can reward/punish it if it does the right/wrong thing.

When watching the video, notice how the program is initially clumsy and unskilled but steadily improves with training until it becomes a champion.

D. Semi-supervised learning:

Where an incomplete training signal is given: a training set with some (often many) of the target outputs missing. There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing. Semi-supervised learning is an approach to machine learning that combines small labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning and supervised learning.

Categorizing based on Required Output

Another categorization of machine-learning tasks arises when one considers the desired output of a machine-learned system:

Classification: When inputs are divided into two or more classes, the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are “spam” and “not spam”.

Regression: Which is also a supervised problem, A case when the outputs are continuous rather than discrete.

Clustering: When a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task.

Machine Learning comes into the picture when problems cannot be solved using typical approaches. ML algorithms combined with new computing technologies promote scalability and improve efficiency. Modern ML models can be used to make predictions ranging from outbreaks of disease to the rise and fall of stocks.

ALGORITHMS

From classification to regression, here are seven algorithms you need to know:

1. Linear regression

Linear regression is a supervised learning algorithm used to predict and forecast values within a continuous range, such as sales numbers or prices.

Originating from statistics, linear regression performs a regression task, which maps a constant slope using an input value (X) with a variable output (Y) to predict a numeric value or quantity.

Linear regression uses labelled data to make predictions by establishing a line of best fit, or 'regression line', that is approximated from a scatter plot of data points. As a result, linear regression is used for predictive modelling rather than categorisation.

2. Logistic regression

Logistic regression, or 'logit regression', is a supervised learning algorithm used for binary classification, such as deciding whether an image fits into one class.

Originating from statistics, logistic regression technically predicts the probability that an input can be categorised into a single primary class. In practice, however, this can be used to group outputs into one of two categories ('the primary class' or 'not the primary class'). This is achieved by creating a range for binary classification, such as any output between 0-.49 is put in one group, and any between .50 and 1.00 is put in another.

As a result, logistic regression in machine learning is typically used for binary categorisation rather than predictive modelling.

3. Naive Bayes

Naive Bayes is a set of supervised learning algorithms used to create predictive models for either binary or multi-classification. Based on Bayes' theorem, Naive Bayes operates on conditional probabilities, which are independent of one another but indicate the likelihood of a classification based on their combined factors.

For example, a programme created to identify plants might use a Naive Bayes algorithm to categorise images based on particular factors, such as perceived size, colour, and shape. While each of these factors is independent, the algorithm would note the likelihood of an object being a particular plant using the combined factors.

4. Decision tree

A decision tree is a supervised learning algorithm used for classification and predictive modelling.

Resembling a graphic flowchart, a decision tree begins with a root node, which asks a specific question of the data and then sends it down a branch depending on the answer. These branches each lead to an internal node, which asks another question of the data before directing it toward another branch, depending on the answer. This continues until the data reaches an end node, also called a leaf node, that doesn't branch any further.

Decision trees are common in machine learning because they can handle complex data sets with relative simplicity.

5. Random forest algorithm

A random forest algorithm uses an ensemble of decision trees for classification and predictive modelling.

In a random forest, many decision trees (sometimes hundreds or even thousands) are each trained using a random sample of the training set (a method known as 'bagging'). Afterwards, the algorithm puts the same data

into each decision tree in the random forest and tallys their end results. The most common result is then selected as the most likely outcome for the data set.

Although they can become complex and require significant time, random forests correct the common problem of 'overfitting' that can occur with decision trees. Overfitting is when an algorithm coheres too closely to its training data set, which can negatively impact its accuracy when introduced to new data later.

6. K-nearest neighbour (KNN) algorithm

A K-nearest neighbour is a supervised learning algorithm for classification and predictive modelling.

True to its name, KNN algorithms classify an output by its proximity to other outputs on a graph. For example, if an output is closest to a cluster of blue points on a graph rather than a cluster of red points, it would be classified as a member of the blue group. This approach means that KNN algorithms can classify known outcomes or predict the value of unknown ones.

7. K-Means algorithm

K-Means is an unsupervised algorithm used for classification and predictive modelling.

Much like KNN, K-Means uses the proximity of an output to a cluster of data points to identify it. Each of the clusters is defined by a centroid, a real or imaginary centre point for the cluster. K-Means is useful on large data sets, especially for clustering, though it can falter when handling outliers.

APPLICATIONS OF MACHINE LEARNING

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:

1. Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, Automatic friend tagging suggestion:

Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's face detection and recognition algorithm.

It is based on the Facebook project named "Deep Face," which is responsible for face recognition and person identification in the picture.

2. Speech Recognition

While using Google, we get an option of "Search by voice," it comes under speech

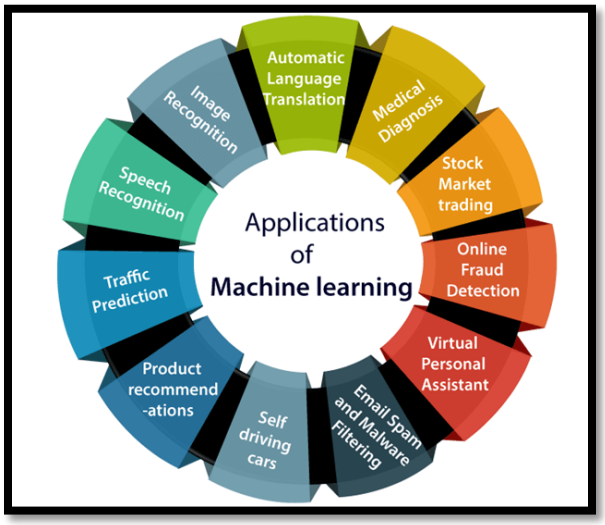

Fig. Applications of Machine Learning

recognition, and it's a popular application of machine learning.

Speech recognition is a process of converting voice instructions into text, and it

is also known as "Speech to text", or "Computer speech recognition." At present, machine learning algorithms are widely used by various applications of speech recognition. Google assistant, Siri, Cortana, and Alexa are using speech recognition technology to follow the voice instructions.

3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

Real Time location of the vehicle form Google Map app and sensors

Average time has taken on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations:

Machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.

As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. Self-driving cars:

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:

- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

Some machine learning algorithms such as Multi-Layer Perceptron, Decision tree, and Naïve Bayes classifier are used for email spam filtering and malware detection.

7. Virtual Personal Assistant:

We have various virtual personal assistants such as Google assistant, Alexa, Cortana, Siri. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.

These virtual assistants use machine learning algorithms as an important part.

These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

8. Online Fraud Detection:

Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as fake accounts, fake ids, and steal money in the middle of a transaction. So to detect this, Feed Forward Neural network helps us by checking whether it is a genuine transaction or a fraud transaction.

For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

9. Stock Market trading:

Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's long short term memory neural network is used for the prediction of stock market trends.

10. Medical Diagnosis:

In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.

It helps in finding brain tumors and other brain-related diseases easily.

7 MAJOR CHALLENGES FACED BY MACHINE LEARNING PROFESSIONALS

1. Poor Quality of Data

Data plays a significant role in the machine learning process. One of the significant issues that machine learning professionals face is the absence of good quality data. Unclean and noisy data can make the whole process extremely exhausting. We don't want our algorithm to make inaccurate or faulty predictions. Hence the quality of data is essential to enhance the output. Therefore, we need to ensure that the process of data preprocessing which includes removing outliers, filtering missing values, and removing unwanted features, is done with the utmost level of perfection.

2. Underfitting of Training Data

This process occurs when data is unable to establish an accurate relationship between input and output variables. It simply means trying to fit in undersized jeans. It signifies the data is too simple to establish a precise relationship. To overcome this issue:

- Maximize the training time
- Enhance the complexity of the model
- Add more features to the data
- Reduce regular parameters
- Increasing the training time of model

3. Overfitting of Training Data

Overfitting refers to a machine learning model trained with a massive amount of data that negatively affect its performance. It is like trying to fit in Oversized jeans. Unfortunately, this is one of the significant issues faced by machine learning professionals. This means that the algorithm is trained with noisy and biased data, which will affect its overall performance. Let's understand this with the help of an example. Let's consider a model trained to differentiate between a cat, a rabbit, a dog, and a tiger. The training data contains 1000 cats, 1000 dogs, 1000 tigers, and 4000 Rabbits. Then there is a considerable probability that it will identify the cat as a rabbit. In this example, we had a vast amount of data, but it was biased; hence the prediction was negatively affected.

We can tackle this issue by:

- Analyzing the data with the utmost level of perfection
- Use data augmentation technique
- Remove outliers in the training set
- Select a model with lesser features

4. Machine Learning is a Complex Process

The machine learning industry is young and is continuously changing. Rapid hit and trial experiments are being carried on. The process is transforming, and hence there are high chances of error which makes the learning complex. It includes analyzing the data, removing data bias, training data, applying complex mathematical calculations, and a lot more. Hence it is a really complicated process which is another big challenge for Machine learning professionals.

5. Lack of Training Data

The most important task you need to do in the machine learning process is to train the data to achieve an accurate output. Less amount training data will produce inaccurate or too biased predictions. Let us understand this with the help of an example. Consider a machine learning algorithm similar to training a child. One day you decided to explain to a child how to distinguish between an apple and a watermelon. You will take an apple and a watermelon and show him the difference between both based on their color, shape, and taste. In this way, soon, he will attain perfection in differentiating between the two. But on the other hand, a machine-learning algorithm needs a lot of data to distinguish. For complex problems, it may even require millions of data to be trained. Therefore we need to ensure that Machine learning algorithms are trained with sufficient amounts of data.

6. Slow Implementation

This is one of the common issues faced by machine learning professionals. The machine learning models are highly efficient in providing accurate results, but it takes a tremendous amount of time. Slow programs, data overload, and excessive requirements usually take a lot of time to provide accurate results. Further, it requires constant monitoring and maintenance to deliver the best output.

7. Imperfections in the Algorithm When Data Grows

So you have found quality data, trained it amazingly, and the predictions are really concise and accurate. Yay, you have learned how to create a machine learning algorithm!! But wait, there is a twist; the model may become useless in the future as data grows. The best model of the present may become inaccurate in the coming Future and require further rearrangement. So you need regular monitoring and maintenance to keep the algorithm working. This is one of the most exhausting issues faced by machine learning professionals.

MACHINE LEARNING TOOLS

Machine learning is one of the most revolutionary technologies that is making lives simpler. It is a subfield of Artificial Intelligence, which analyses the data, build the model, and make predictions. Due to its popularity and great applications, every tech enthusiast wants to learn and build new machine learning Apps. However, to build ML models, it is important to master machine learning tools. Mastering machine learning tools will enable you to play with

the data, train your models, discover new methods, and create algorithms.

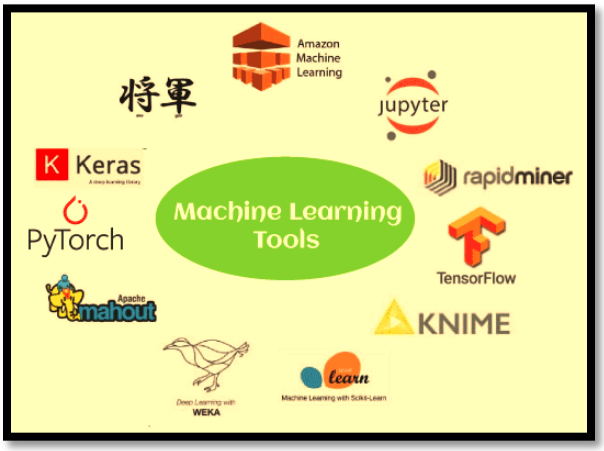

Fig. Machine Learning Tools

1. TensorFlow

TensorFlow is one of the most popular open-source libraries used to train and build both machine learning and deep learning models. It provides a JS library and was developed by Google Brain Team. It is much popular among machine learning enthusiasts, and they use it for building different ML applications. It offers a powerful library, tools, and resources for numerical computation, specifically for large scale machine learning and deep learning projects. It enables data scientists/ML developers to build and deploy machine learning applications efficiently. For training and building the ML models, TensorFlow provides a high-level Keras API, which lets users easily start with TensorFlow and machine learning.

2. PyTorch

PyTorch is an open-source machine learning framework, which is based on the Torch library. This framework is free and open-source and developed by FAIR(Facebook's AI Research lab). It is one of the popular ML frameworks, which can be used for various applications, including computer vision and natural language processing. PyTorch has Python and C++ interfaces; however, the Python interface is more interactive. Different deep learning software is made up on top of PyTorch, such as PyTorch Lightning, Hugging Face's Transformers, Tesla autopilot, etc.

It specifies a Tensor class containing an n-dimensional array that can perform tensor computations along with GPU support.

3. Google Cloud ML Engine

While training a classifier with a huge amount of data, a computer system might not perform well. However, various machine learning or deep learning projects requires millions or billions of training datasets. Or the algorithm that is being used is taking a long time for execution. In such a case, one should go for the Google Cloud ML Engine. It is a hosted platform where ML developers and data scientists build and run optimum quality machine, learning models. It provides a managed service that allows developers to easily create ML models with any type of data and of any size.

4. Amazon Machine Learning (AML)

Amazon provides a great number of machine learning tools, and one of them is Amazon Machine Learning or AML. Amazon Machine Learning (AML) is a cloud-based and robust machine learning software application, which is widely used for building machine learning models and making predictions. Moreover, it integrates data from multiple sources, including Redshift, Amazon S3, or RDS.

FUTURE SCOPE OF MACHINE LEARNING

The scope of Machine Learning is not limited to the investment sector. Rather, it is expanding across all fields such as banking and finance, information technology, media & entertainment, gaming, and the automotive industry. As the Machine Learning scope is very high, there are some areas where researchers are working toward revolutionizing the world for the future. Let us discuss them in detail.

CONCLUSION

The voyage to master machine learning is an exhilarating and ever-evolving expedition that transcends the boundaries of data and algorithms, forging a path toward transformative innovation. Throughout this comprehensive exploration, we've traversed the foundational bedrock of machine learning, unraveling its fundamental concepts, algorithms, and diverse learning pathways.

This journey is not merely a trajectory of academic pursuit; it is an immersive experience encapsulating practical applications across multifarious domains. From healthcare to finance, marketing to robotics, machine learning stands as an omnipotent catalyst, enabling solutions to intricate real-world challenges.

REFERENCES

1. ChatGPT
2. GeeksForGeeks
3. Intellipat
4. JavatPoint