IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Artificial Intelligence and Machine learning

Dr. Ashima Metha[1], Akib Khan[2], Uday Sakhare[3]
Mentor, Department of Computer Science Engineering[1]
Students, Department of Computer Science Engineering[2,3]
Dronacharya College of Engineering, Gurugram , India

Abstract: *This article illustrates the applicability of AIML in the management processes of knowledge bases, with particular attention to applicability in the contexts of e-learning. For this purpose, the authors discuss a Tutor Bot planned to be able to support the learners of an e-learning platform; moreover, Tutor Bot has the possibility, aside from manipulating direct questions in natural language, to produce research on the diffused knowledge bases (e.g., the Internet) through direct*
interaction with a search engine. Finally, the management problem in AIML of retrieving information of Tutor Bot is faced in the interaction with the search engine; to this purpose, the authors suggest a new tag used in AIML that allows for management of this specificity.

Keywords: AIML

I. INTRODUCTION

Data Mining:
Data mining is the process of extracting and discovering patterns in large datasets involving methods at the intersection of machine learning, statistics,
and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal of extracting information (with intelligent methods) from a data set and transforming the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Aside from the raw analysis step, it alsoinvolves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discoveredstructures, visualization, and online updating
The term "data mining" is a misnomer because the goal is the extraction
of patterns and knowledge from large amounts of data, not the extraction (*mining*) of data itself. Often the more general terms (*large scale*) *data analysis* and *analytics*— or, when referring to actual methods, *artificial intelligence* and *machine learning*—aremore appropriate.
The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groupsof data records (cluster analysis), unusual records (anomaly detection),
and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can thenbe seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, although they do belong to the overall KDD process as additional steps.

II. MACHINE LEARNING

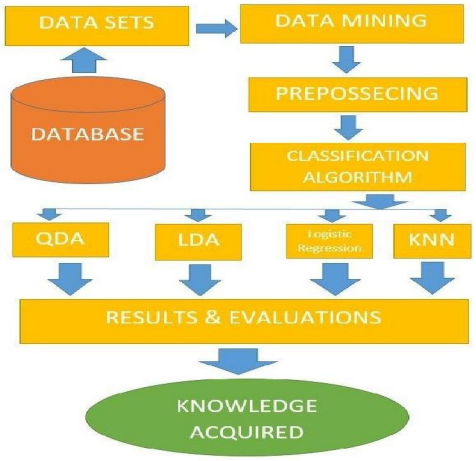
Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. Recently, artificial neural networks have been able to surpassmany previous approaches in performance.
Machine learning approaches have been applied to many fields including naturallanguage processing, computer vision, speech recognition, email filtering, agriculture, and medicine. ML is known in its application across businessproblems

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

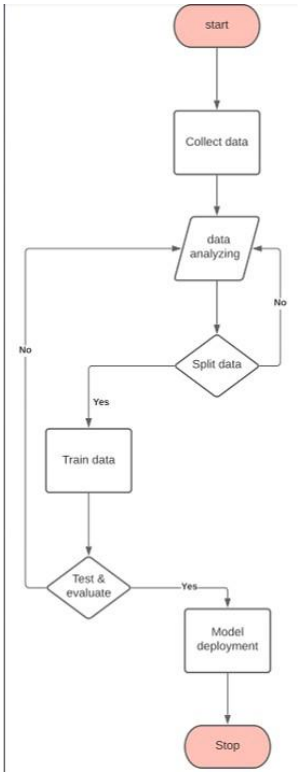
under the name predictive analytics. Although not all machine learning is statistically based, computational statistics is an important source of the field's methods.

The mathematical foundations of ML are provided by mathematical optimization (mathematical programming) methods. Data mining is a related(parallel) field of study, focusing on exploratory data analysis (EDA) through unsupervised learning.

From a theoretical viewpoint, probably approximately correct (PAC) learning provides a framework for describing machine learning



Flow chart of Data mining



Machine learning flow chart

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Relationship of Machine learning to field of Data Mining

Machine learning and data mining often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on *known* properties learned from the training data, data mining focuses on the discovery of (previously) *unknown* properties in the data (this is the analysis step of knowledge discovery in databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy. Much of the confusion between these two research communities (which do often have separate conferences and separate journals, ECML PKDD being a major exception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability

to *reproduce known* knowledge, while in knowledge discovery and data mining (KDD) the key task is the discovery of previously *unknown* knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the unavailability of training data.

Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set of examples).

	Data Mining	Machine Learning
Focus	Discovery of hidden pattern	Development of algorithm that learn from data
Goal	Extract insight and information from existing datasets	Build models to make predictions
Usage	Identifying patterns	Predictive modelling, classification, lustring etc
Input	Historical data or large datasets	Labelled or unlabelled data for training and testing
Output	Knowledge in the form of patterns	Predictions, classification, recommendation etc
Scope	Broder in terms of analysingvarious types of data	Focused on development models for specific application
Domain	Widely used in business, marketing, healthcare etc	Widely used in AI, robotics, pattern etc.

Comparison of Data mining and Machine learning

Implementation of Data mining

Data mining is described as a process of finding hidden precious data by evaluatingthe huge quantity of information stored in data warehouses, using multiple data mining techniques such as Artificial Intelligence (AI), Machine learning and statistics.Data mining implementation process consists of 6 steps:

Business understanding:

It focuses on understanding the project goals and requirements form a business point of view, then converting this information into a data mining problem afterwarda preliminary plan designed to accomplish the target.

Tasks:

- Determine business objectives
- Access situation
- Determine data mining goals

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

- Produce a project plan

Data Understanding:
Data understanding starts with an original data collection and proceeds with operations to get familiar with the data, to data quality issues, to find better insightin data, or to detect interesting subsets for concealed information hypothesis.
Tasks:

- Collects initial data
- Describe data
- Explore data
- Verify data quality

Data Preparation
It takes more time. Covers all operations to build the final data set from the originalraw information. Several times, data preparation is probable to be done
Tasks:

- Select data
- Clean data
- Construct data
- Integrate data
- Format data

Modeling:
In modeling, various modeling methods are selected and applied, and their parameters are measured to optimum values. Some methods gave particular requirements on the form of data. Therefore, stepping back to the data preparationphase is necessary.
Tasks:

- Select modeling technique
- Generate test design
- Build model
- Access model

Evaluation:
At the last of this phase, a decision on the use of the data mining results should be reached. It evaluates the model efficiently, and review the steps executed to build the model and to ensure that the business objectives are properly achieved. The main objective of the evaluation is to determine some significant business issue that has not been regarded adequately. At the last of this phase, a decision on the use of the data mining outcomes should be reached.
Tasks:

- Evaluate results
- Review process
- Determine next steps

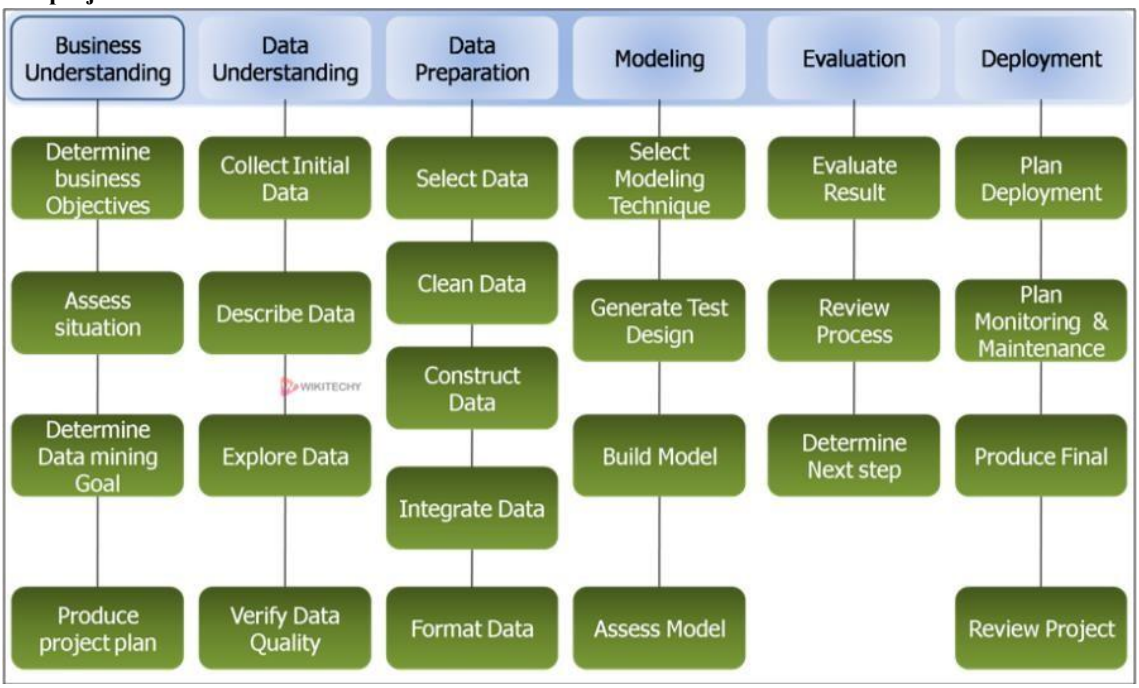
Deployment
The concept of deployment in data mining refers to the appliance of a model for prediction using a new data. The deployment phase are often as simple as generating a report or as complex as implementing a repeatable data mining process.
Tasks

- Plan deployment
- Plan monitoring and maintenance

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

- Produce final report

Review project



Implementation of Machine Learning

Machine learning has become an essential part of modern software development, enabling applications to learn from data and make intelligent decisions.

Incorporating machine learning algorithms into your software projects can lead to more personalized user experiences, improved efficiency, and better decision-makingcapabilities. In this step-by-step guide, we will walk you through the process of implementing a basic machine learning algorithm in your software development project.

Step 1: Define the Problem and Gather Data

The first step in any machine learning project is to clearly define the problem you want to solve. Whether it's predicting customer churn, classifying images, or recommending products, a well-defined problem statement is crucial. Consider thebusiness objectives and the data you have or need to collect.

Once you have a clear problem statement, gather relevant data from various sources.Ensure the data is clean, free from errors, and representative of the problem you are trying to solve. Data quality is essential for the success of your machine learning model.

Step 2: Preprocess and Explore the Data

Before feeding the data into a machine learning model, it requires preprocessing. Common preprocessing steps include handling missing values, scaling features, andconverting categorical variables to numerical representations.

Explore the data to gain insights and understand the relationships between differentfeatures. Data visualization tools can be beneficial in this step to identify patterns, correlations, and potential outliers.

Step 3: Split the Data into Training and Testing Sets

To evaluate the performance of your machine learning model, split the data into training and testing sets. The training set will be used to train the model, while the testing set will assess its performance on unseen data. A typical split ratio is around80/20 or 70/30 for training and testing sets, respectively.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Step 4: Choose a Basic Machine Learning Algorithm
Select a basic machine learning algorithm appropriate for your problem. For beginners, linear regression for regression tasks and logistic regression for classification tasks are good starting points. These algorithms are relatively easy to understand and implement.

Step 5: Train the Model
With the training data ready, you can now train the machine learning model using the chosen algorithm. During the training process, the model will learn from the dataand adjust its internal parameters to make accurate predictions.

Step 6: Evaluate the Model
After the model has been trained, it's time to evaluate its performance on the testingdata. Common evaluation metrics include accuracy, precision, recall, F1 score, and mean squared error, depending on the nature of the problem

Step 7: Tune Hyperparameters (Optional)
Many machine learning algorithms have hyperparameters that affect the model'sperformance. Hyperparameters are set before training the model and can significantly impact its accuracy. Experiment with different values for these hyperparameters to find the best combination that maximizes the model's performance.

Step 8: Deploy the Model
Once you are satisfied with the model's performance, it's time to deploy it in your software development project. Depending on your application, you can deploy the model as a part of your software's backend or use it as a standalone service throughAPIs.

Step 9: Monitor and Update
Machine learning models are not static; they need regular monitoring and updates.As your software gathers more data, the model may need to be retrained with the new data to maintain its accuracy and relevancy.
Incorporating machine learning algorithms into your software development projects can add tremendous value and empower your applications to make intelligent decisions. By following this step-by-step guide, you can implement a basic machine learning algorithm successfully. Remember that machine learning is an iterative process, so don't be afraid to experiment and continuously improve your models.
With time and experience, you'll be able to tackle more complex machine learningchallenges and further enhance your software projects. Happy coding!