Artificial Intelligence/ Machine Learning/ Data Analytics

About:

Artificial Intelligence is everywhere, from gaming stations to maintaining complex information at work. Computer Engineers and Scientists are working hard to impart intelligent behavior in the machines making them think and respond to real-time situations. AI is transiting from just a research topic to the early stages of enterprise adoption. Tech giants like Google and Facebook have placed huge bets on Artificial Intelligence and Machine Learning and are already using it in their products. But this is just the beginning, over the next few years, we may see AI steadily glide into one product after another.

According to Stanford Researcher, John McCarthy, "*Artificial Intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs. Artificial Intelligence is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.*"

Simply put, AI's goal is to make computers/computer programs smart enough to imitate the human mind's behaviour.

Artificial Intelligence and Machine Learning are much trending and also confusing terms nowadays. Machine Learning (ML) is a subset of Artificial Intelligence. ML is a science of designing and applying algorithms that are able to learn things from past cases. If some behaviour exists in the past, then you may predict if it can happen again. Means if there are no past cases then there is no prediction.

ML can be applied to solve tough issues like credit card fraud detection, enable self-driving cars and face detection and recognition. ML uses complex algorithms that constantly iterate over large data sets, analyzing the patterns in data and facilitating machines to respond to different situations for which they have not been explicitly programmed. The machines learn from history to produce reliable results. The ML algorithms use Computer Science and Statistics to predict rational outputs.

The term data analytics refers to the process of examining datasets to draw conclusions about the information they contain. Data analytic techniques enable you to take raw data and uncover patterns to extract valuable insights from it.

Today, many data analytics techniques use specialized systems and software that integrate machine learning algorithms, automation and other capabilities.

Data has the potential to provide a lot of value to businesses, but to unlock that value, you need the analytics component. Analysis techniques give businesses access to insights that can help them to improve their performance. It can help you improve your knowledge of your customers, ad campaigns, budget and more.

As the importance of data analytics in the business world increases, it becomes more critical that your company understand how to implement it.

With this background, we have grouped these THREE topics under one domain. Since they are tightly coupled with each other, we can always explore them in general to begin with and later, focus on the particular topic based on the interest.

Pre-Requisites:

Strong foundation in Math

Calculus

Algebra

Probability

Statistics

Topics Involved:

- Mathematics & Statistics

  https://www.freecodecamp.org/news/statistics-for-data-science/

  https://www.youtube.com/watch?v=o8f2w0Q0ME4

- MS Excel

  https://www.youtube.com/watch?v=xhFDYKqWlqw

- Python Programming

  https://www.youtube.com/watch?v=rfscVS0vtbw

- Pandas and Numpy

https://www.youtube.com/watch?v=GPVsHOlRBBI

- TensorFlow

https://www.youtube.com/watch?v=tPYj3fFJGjk

- ML Models
- Matlab/Octave
- Power BI
- Tableau
- IBM Watson
- Jupyter

https://www.youtube.com/watch?v=HW29067qVWk

- skLearn

https://www.youtube.com/watch?v=pqNCD_5r0IU

Syllabus:

Level 1:

1. Linear Regression vs Logistic Regression
   a) Predict the price of a home based on multiple different variables using Boston House Dataset (Linear Regression)
   b) Train a model to distinguish between different species of the Iris flower based on sepal length, sepal width, petal length, and petal width (Logistic Regression)

   https://www.javatpoint.com/linear-regression-vs-logistic-regression-in-machine-learning#:~:text=Linear%20regression%20is%20used%20to,given%20set%20of%20independent%20variables.&text=Logistic%20regression%20is%20used%20for%20solving%20Classification%20problems.

2. Data Visualisation using matplotlib and diff kinds of plots. (Bar, scatter, Time Series and Histogram)

3. Implement K- Nearest Neighbour Algorithm for Iris dataset

https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e#:~:text=What%20is%20KNN%3F,how%20its%20neighbours%20are%20classified

4. Implement Decision tree based ID3 algorithm using appropriate dataset and apply this to classify a new sample https://iq.opengenus.org/id3-algorithm/

5. Implement Naive Bayesian Classifier for text classification. Naive Bayes theorem is based on probabilities of events https://www.geeksforgeeks.org/naive-bayes-classifiers/

6. Ensemble techniques combine the decisions from multiple models to improve the overall performance. Apply the ensemble techniques on the Titanic Dataset, https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

7. Exploratory Data analysis on Airbnb Data Formulate the scraped data into features that will assist the model to predict the listing's price using exploratory data analysis on the dataset

   https://towardsdatascience.com/exploratory-data-analysis-feature-engineering-and-modelling-using-supermarket-sales-data-part-1-228140f89298

8. Hyperparameter Tuning is used to fine-tune a model in order to obtain higher accuracy. Build a cat vs dog classifier and hyper tune in the parameters to obtain higher frequency model.

   https://www.geeksforgeeks.org/hyperparameter-tuning/

9. Image Classification using KMeans Clustering Classify a given set of images into a given number of categories using KMeans Clustering using MNIST dataset.

   https://towardsdatascience.com/image-clustering-using-k-means-4a78478d2b83

10. Breast Cancer Classification with the help of Support Vector Machines - Correct and timely diagnosis of Breast Cancer, an exquisite disease is an essential matter in the medical field. Using the concept of Support Vector Machines, detect the possibility of breast cancer.

https://www.freecodecamp.org/news/svm-machine-learning-tutorial-what-is-the-support-vector-machine-algorithm-explained-with-code-examples/

11. Building a Machine Learning Model using H20.ai. Build a model and train the predictions using H20 framework (For eg. predicting quality of wine)

    https://www.kdnuggets.com/2020/01/h2o-framework-machine-learning.html

12. Build an Artificial Neural Network using Back Propogation Algorithm.

    https://www.ibm.com/cloud/learn/neural-networks

Level 2:

A Bank wants to take care of customer retention for their product: savings accounts. The bank wants you to identify customers likely to churn balances below the minimum balance in the next quarter. You have the customers information such as age, gender, demographics along with their transactions with the bank. Your task as a data scientist would be to predict the propensity to churn for each customer.

· Uber, Lyft, Ola and many more online ride hailing services are trying hard to use their extensive data to create data products such as pricing engines, driver allotment etc. To improve the efficiency of taxi dispatching systems for such services, it is important to be able to predict how long a driver will have his taxi occupied or in other words the trip duration. This project will cover techniques to extract important features and accurately predict trip duration for taxi trips in New York using data from TLC commission New York.

· Leveraging IBM Watson's Natural Language Processing capabilities, you'll learn how to plan, implement, test, and deploy chatbots

Level 3:

You now have the skills required to work on an AI/ML project of your own. Feel free to use any/all of the skills and equipment you have learnt in the last 2 levels.

Assessment:

Level 1:

After completing this level, upload a photograph/video of the functioning circuit or model and show it to your domain co-ordinator.

You are required to provide a detailed report consisting of:

1. Project specifications
2. Challenges faced
3. Documentation (comments in your the code)
4. Material used
5. What you have learnt

After successfully completing the report, you'll be allowed to move on to level 2.

Level 2:

After completing this level, upload a photograph/video of the functioning circuit or model and show it to your domain co-ordinator.

You are required to provide a detailed report consisting of:

1. Project specifications
2. Challenges faced
3. Documentation
4. Material used

5. What you have learnt

A date will be fixed on which you give a brief seminar on your report. After successfully completing both the report and seminar, you'll be allowed to move on to level 2.

Level 3:

Level 3 assessment will be on a case-by-case basis. Contact your coordinator for more details.

References:

- https://github.com/abhishekkrthakur/approachingalmost/blob/master/AAAMLP.pdf
- https://builtin.com/artificial-intelligence
- https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning
- https://cloud.ibm.com/developer/watson/documentation

Useful Links:

- https://docs.microsoft.com/en-us/windows/ai/windows-ml/
- https://www.appliedaicourse.com/course/11/Applied-Machine-learning-course
- https://aditya-bhattacharya.net/blogs/
- https://matplotlib.org/
- https://www.freecodecamp.org/news/how-to-build-and-train-k-nearest-neighbors-ml-models-in-python/

Courses:

- https://www.coursera.org/learn/machine-learning
- https://www.coursera.org/learn/ai-for-everyone
- https://www.coursera.org/specializations/tensorflow-data-and-deployment
- https://www.coursera.org/specializations/mathematics-machine-learning
- https://www.coursera.org/professional-certificates/google-data-analytics
- https://www.coursera.org/learn/ai-with-ibm-watson
- https://www.edx.org/course/machine-learning-with-python-a-practical-introduct
- https://www.edx.org/course/cs50s-introduction-to-artificial-intelligence-with-python
- https://www.edx.org/course/pytorch-basics-for-machine-learning
- https://www.coursera.org/specializations/excel-data-analytics-visualization
- https://www.coursera.org/specializations/data-analysis-visualization-foundations
- https://www.coursera.org/specializations/excel

- https://www.coursera.org/learn/stanford-statistics

- https://www.coursera.org/learn/excel-basics-data-analysis-ibm

- https://ai.google/education/

Notes:

https://drive.google.com/file/d/1jg5c7ZVwEahWtCfGQ0hOdKsDHR1oulrc/view?usp=sharing