The Evolution of Machine Learning

1. Introduction and background

Machine learning. Machine learning (ML) is a critical technique for utilizing artificial intelligence-related technologies. Machine learning is frequently referred to as AI due to its learning and decision-making capabilities, although it is actually a subset of AI [1]. It was a natural progression of AI until the late 1970s [3]. Then it evolved independently. Machine learning has developed into a critical response tool for cloud computing and e-Commerce and is now being used to a range of cutting-edge technologies. For many firms today, machine learning is a required component of contemporary business and research. It assists computer systems in gradually improve their performance by utilizing algorithms and neural network models [1]. Machine learning algorithms create a mathematical model automatically using sample data – often called "training data" – in order to make judgments without being explicitly programmed to do so [1].

2. Objectives and goals

In this assignment, I want to analyze how ML techniques evolve over time and which fields are utilizing ML techniques. I want to also analyze how well ML is doing in different fields. Furthermore, I want to observe what is the future of ML and what methods are currently developed by the ML community and how these methods are used in different applications. I will demonstrate above concerns by visualizing the following questions,

1. How machine learning evolves over time?
2. Does the community from various fields accepted machine learning for their application?
3. Which countries are doing well in machine learning research?

These analyses will help what are state-of-the-art methods for ML. Which areas adopted ML-based methods and how well ML is performing over the world. For, instance by observing the cooccurrence of the keywords in the ML related papers one can understand which are the current hot topics for ML, decide the future research, and write research proposal for funding and use this information as supporting evidence. Additionally, we can also evaluate which countries are doing well in ML research by observing the citation counts on the ML papers published over the world. Finally, the analysis might be also helpful for different industries to determine whether they utilize ML for their applications.

3. Datasets

I used Web of Science which is accessible via Thomas Cooper Library and used Web of Science Core Collection to collect the machine learning papers published between 2010-2022 [2]. The dataset is available at:
https://github.com/abirhossen786/visualization_tools_coursework/tree/main/iviz3/dataset.
The dataset contains 3000 records of full publication details.

Fields include
- id
- label
- x
- y
- cluster weight<Links>
- weight<Total link strength>
- weight<Occurrences>
- score<Avg. pub. year>
- score<Avg. citations>
- score<Avg. norm. citations>
- Author name
- Country

4. Visualization Plan

How the applications of machine learning evolve over time?

To answer the question, I choose co-word analysis. I mainly focused on occurrences of keywords by authors between 2010-2022. I think this specific visualization is best because it shows the depth of the occurrences. As shown in Figure 1, Figure 2, and Figure 3, during 2010-2012 ML just started to evolve, support vector machine [4], random forest [5] methods are applied for classification tasks. We also see, researcher just started applying extreme learning machine [6]. Additionally, ML is also applied in scheduling task. We can conclude, during 2010-2012 the research community just started applying ML for simple classification tasks
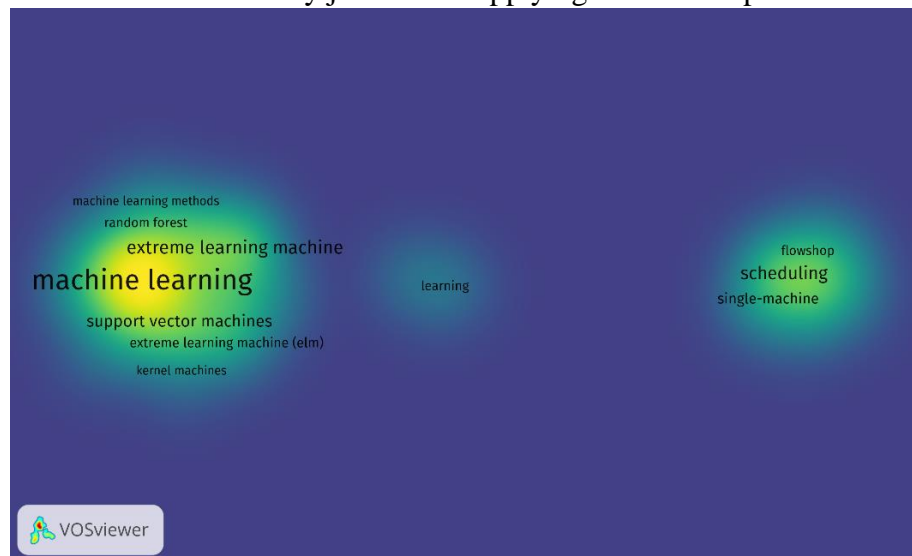

Figure 1. Occurrences of keywords by authors between 2010-2012

During 2015-2017, we see a lot more areas where ML was applied including big data, natural language processing, feature learning, bioinformatics. We also observe some new terms such as, deep learning, neural networks, optimization, supervised learning, active learning just

introduced during this time. Extreme learning machine also gain popularity compared to 2010-2012.
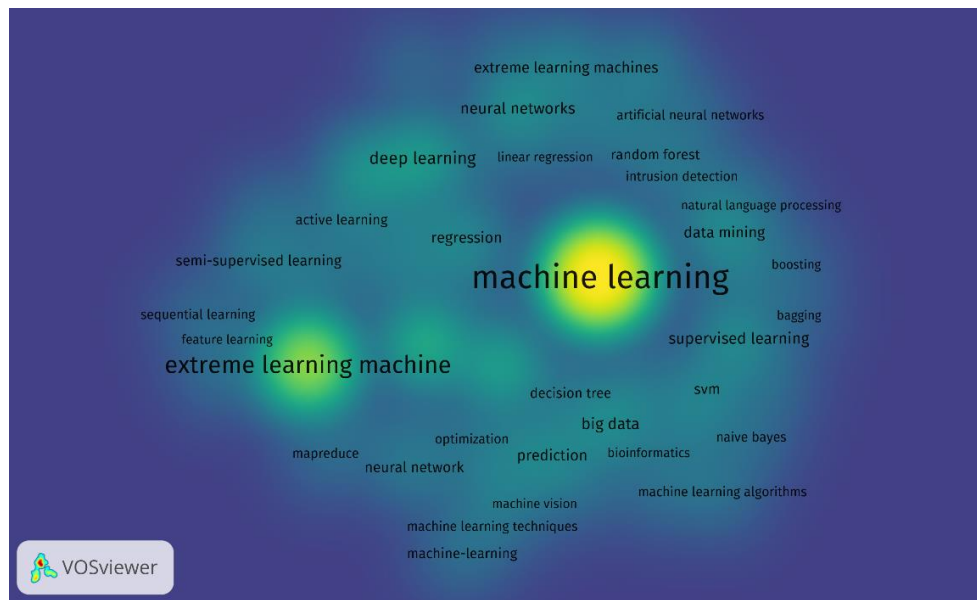

Figure 2. Occurrences of keywords by authors between 2015-2017

During 2020-2022, we see deep learning gained huge attention by the community and shows researchers adopted deep learning and AI techniques. We also observe more diverse application of ML. In addition, some new methods such as untreatable ML, explainable ML draw attention and research are being conducted in these areas. In collusion, we can say ML received significant attention between 2015-2022 in various fields and widely accepted that shows the how powerful tool it is.
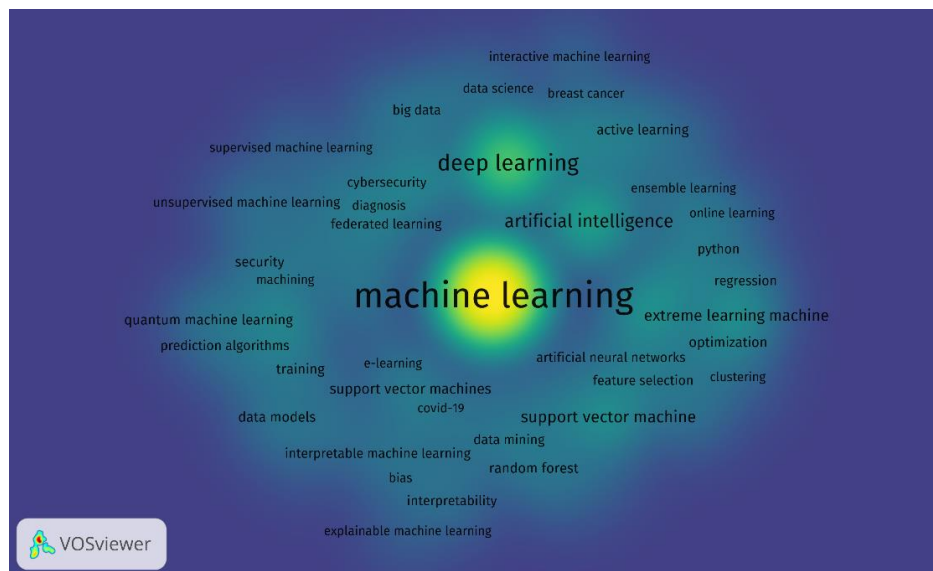

Figure 3. Occurrences of keywords by authors between 2020-2022

Does the community from various fields accepted machine learning for their application?
To answer the question, I choose Network analysis visualization. I mainly focused on link strength of machine learning applied in different applications. I think this specific visualization is best because it shows connection between ML and the filed it was applied. I divided the years into three categories 1) When the ML algorithms were proposed (2010-2012); 2) When ML algorithms established and applied in a few different fields (2015-2017); and 3) Extensive use of ML (2020-2022).

When the ML algorithms were proposed (2010-2012). During this time, people just started developing and testing the ML techniques. Except scheduling and setup times we don't see any use of ML. It was mostly applied on classification tasks.
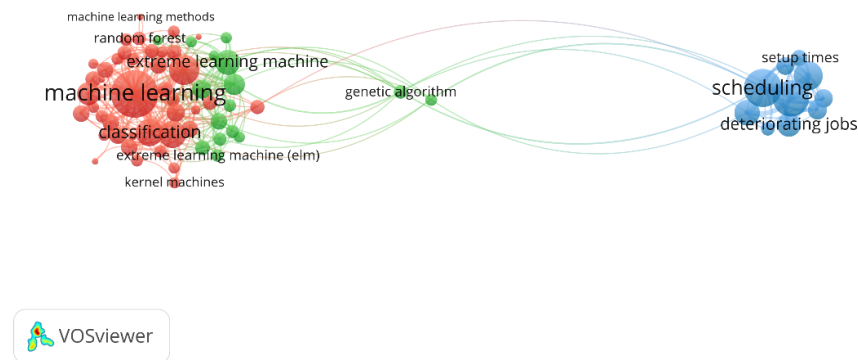

Figure 4. Use of ML between 2010-2012

When ML algorithms established and applied in a few different fields (2015-2017). During this duration, we see more new techniques of ML such as deep learning, and neural networks. We see a lot more areas where ML was applied including big data, data mining, bioinformatics, machine vision. Also, most ML methods were developed during this timeline such as, supervised, and semi-supervised learning, active learning.
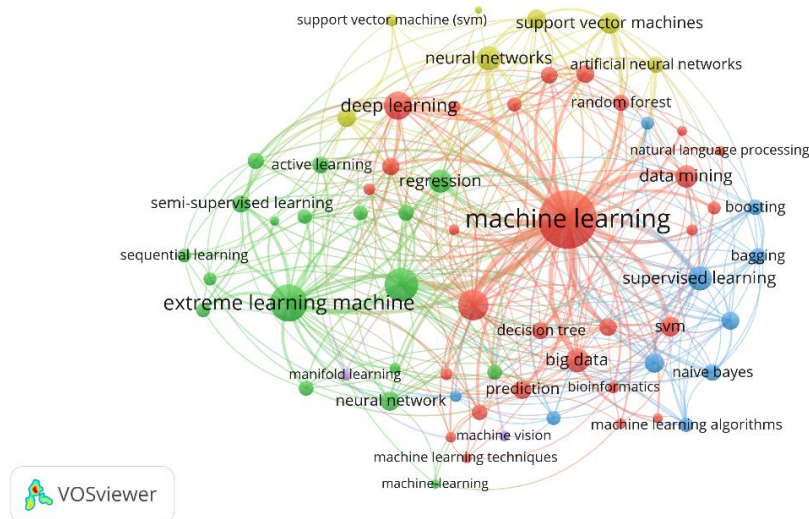

Figure 5. Use of ML between 2015-2017

Extensive use of ML (2020-2022). During this we a significant use of ML across the fields including image processing, materials science, materials informatics, healthcare, feature extractions, quantum computing, social media, genomics We also, we some new methods such as explainable ML, ML optimization, transfer learning, LSTM. We can predict that in next few years we will see a lot more applications using these methods. We can conclude that, the community across the fields accepted ML.
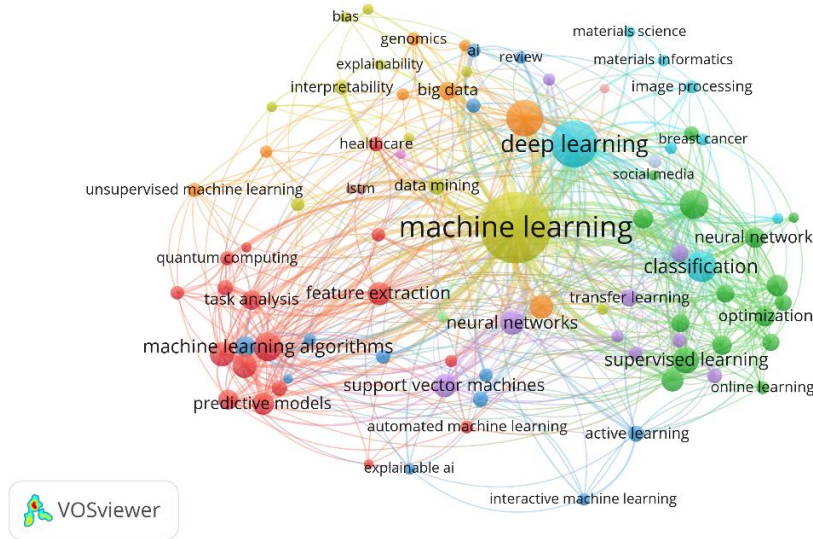

Figure 6. Use of ML between 2020-2022

Which countries are doing well in machine learning research?
To answer this question, I used citation count as an evaluation metric and use link strength of the network analysis methods. We see that USA, Germany, France, Japan, England, and China are ahead compared to other countries. Surprisingly Canada has significantly lower citations count even after residing beside USA, find the reason would be a very interesting research topic.
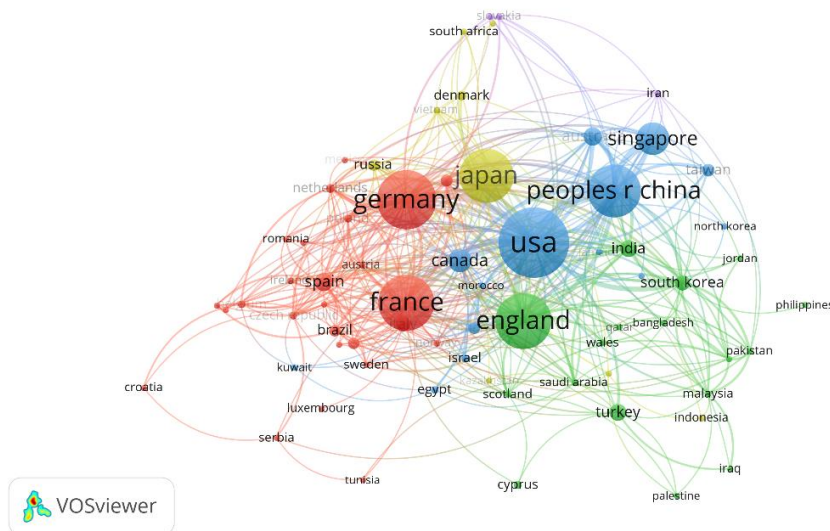

Figure 7. Citations of countries between 2010-2022

5. References

[1] A Brief History of Machine Learning - DATAVERSITY. (2022). Retrieved 17 April 2022, from *https://www.dataversity.net/a-brief-history-of-machine-learning/*

[2] Proxy Login - University Libraries - USC. (2022). Retrieved 17 April 2022, from https://www-webofscience-com.pallas2.tcl.sc.edu/wos/woscc/summary/b5c5d3ef-4e56-4b34-b94d-5776dcd34ff7-319385f4/relevance/1

[3] When Was Machine Learning Invented?. (2022). Retrieved 17 April 2022, from https://pandio.com/blog/when-was-machine-learning-invented/

[4] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[5] Qi, Y. (2012). Random forest for bioinformatics. In Ensemble machine learning (pp. 307-323). Springer, Boston, MA.

[6] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.

Appendix

The dataset and the source VOS viewer files of the above plots are available at:
https://github.com/abirhossen786/visualization_tools_coursework/tree/main/iviz3