Mohammad Ausaf
Literature Review of Research Papers in the Machine Learning Domain

1)An Introduction to Machine Learning

By
Solveig Badillo, Balazs Bandai, Fabian Birzele, lakov I.Davydov, Lucy Hutchinson.

- There have been huge advancements in the field of pharmacometrics and clinical pharmacology as having advancements in Artificial Intelligence and Machine Learning, this paper is supposed to give people in medical sciences an introduction to Machine Learning.

- This paper deals with the mechanisms and procedures by which Machine learning works, and its algorithms.

- In Data and Features section authors discuss what the input data is in the case of clinical settings for instance disease history, blood test results and gene makeup.

- They now move on to explain learning methods, that as Unsupervised Learning, in particular, they discuss the clustering of data, and how it is relevant in clinical practice.

- The authors discuss unsupervised learning methods, identifying what data can be "useful" or can act as training data to obtain output values.

- In the Discussion section, the authors put forth their final thoughts on how AI and ML in their entirety can be used in advancements in patient treatment and clinical practices as predictive methods.

- Conclusions derived: In the age of big data, there are many new opportunities for ML in clinical pharmacology. For example, data generated from wearable devices pose new challenges on how they can be linked to predictive modelling in the future. In addition, access to real-world data could provide strong evidence for covariates, supplement control datasets, and bolster models that have been trained on small datasets.

- The clinical pharmacology community will continue to base their analyses on pharmacological principles and will gradually build new ML elements to their workflow, strengthening their models further. In addition, the clinical pharmacology community will be able to enhance the range of questions they can address by using ML approaches.

2) Certification systems for machine learning: Lessons from sustainability

By:
 Kira J.M. Matus, Michael Veale
Published in Regulation & Governance 2021

Via: Online Wiley Library.

- This paper discusses the need for standardized certification systems in the area of application as well as the development of Artificial Intelligence. Though some regulations exist in the forms of EU's AI acts and some similar but highly insufficient acts and legislations.
- The author discusses how these issues can be analyzed using information about software objects, and information about its deployment context and downstream effect.
- The author discusses how private governance can be the key to solving machine learning's adverse footprints in the environment.
- In the last section, the author discusses how we can learn from pre-existing Sustainability certification systems.
- In conclusion, the author reiterates the importance of measurement and regulations of data processing and the likes of processes in carbon footprints.
- The method or framework proposed by the author is divided into three key parameters, being, standard setting, behaviour modification and information collection, to be elaborate,
    - A set of *technical standards* (which can cover management systems, methods of production, outcomes, or some combination therein). These standards answer the question of *what* is being regulated (what behaviour needs to be modified).
    - A *certification process* (which includes monitoring and enforcement) to ensure that the standards are being met/followed. The certification process answers the question of *how* it will be regulated – specific mechanisms for monitoring and enforcement, including appropriate auditors, who are responsible for information collection.
    - A *labelling program* (to provide information on the credence qualities to consumers). This answers the question of *communication* of participation to the "market".

3) Machine Learning Algorithms for teaching AI to chatbots,

By: Evgeny Tebenkov, Igor Prokhorov
Published in Procedia Computer Science 2021

Via: Elsevier.

- This paper discusses chatbots and their training in particular. Chatbots have become widespread in messengers and social networks and due to the proliferation of chatbot platforms have become easy to create and use.

- The author tells how although there are abundant of these chatbots, bots capable of emulating a human, that is, bots with artificial intelligence are still rare, this is due to the lengthy and complicated imperfect training algorithms.

- In the paper, the author analyzes and compares between machine learning method from a theoretical perspective, and an implementation of a chatbot using NLU (Natural Language Understanding) is given.

- The author provides different machine learning algorithms under the groups of Supervised, Unsupervised, and semi-supervised machine learning algorithms, on the bases of their learning style. Further, the authors give advanced and specific algorithms namely, Neural Networks, Decision Tree Algorithms, Bayesian Algorithms, Regression algorithms, Support Vector, and a few more including ANNs and Clustering algorithms.

- Implementation of Natural Processing Engines or otherwise known as NLP, here the author extracts meaning from natural language to determine the function of a text/sentence, for instance, the word "Please" in a sentence indicates, the function=request, this is called Dialogue Act Recognition.

- Going forward, the authors explore Bayesian approaches to DA (Dialogue act) models, the idea is to find the possibility of every sequence of dialogue acts that could represent a sentence or utterance, and thus fund the DA model with the highest probability of occurring.

- The authors then show an implementation of the above approach in a Facebook messenger chatbot interacting with a real user. In an admittedly small focus group of 20 people, around 12 (60%) were satisfied with the answers and convinced of the bots' requests, while 5 (25%) were satisfied with the answers but refused the bots' requests.

- The article considered 10 machine learning algorithms that can be described for training chatbots, disclosed the IT architecture of a chatbot platform using the NLU Engine learning Bayesian method and the results of testing this chatbot with real users are given. Each of the algorithms described in the article has its advantages and disadvantages, but if used correctly, they allow you to complete any task assigned to the chatbot.

4) Online Social Network Security: A Comparative Review Using Machine Learning and Deep Learning

By Chanchal Kumar, Taran Singh Bharati, Shiv Prakash

- This paper essentially deals with the methods to prevent and recover from various classes of cyber security attacks or threats. This paper comprehensively surveys the evolution of online social networks, their associated risks, and solutions.

- The various security models and state-of-the-art algorithms have been discussed along with a comparative meta-analysis using machine learning, deep learning, and statistical testing to recommend a better solution.

- The authors discuss types of security threats on a social network broadly classified into two categories
    - The First is Risk related to the organization: The threat to the software/application that which organization use for personal or official purposes, any severe attack on any such application may put the network of the entire organization at risk.
    - The second is Risk related to the people: Often than not, people intentionally or accidentally reveal or expose their personal information on their social network.

- The author to achieve security of the said OSN(Online Social Networks) recognize the problem, which is the voluminous data, this huge amount of data makes it incredibly hard to make sense of.

- The authors propose to use High-performance computers, essentially supercomputers to process the data in highly specialized parallel computing algorithms.

- The Quality of service (QoS) is measured on the following parameters, Accuracy (A), Precision (P), Recall (R), and F score (F). The authors then go on to compare the results of 9 different algorithms including SVM, KNN, ANN, AIS, AIS(Artificial Immune System), RS(Rough Sets), RVM(Relevance vector Machines), RF (Random Forest), LR(Logistic Regression).

- Metrics used in the comparative analysis were True positives, True Negatives, False Positives, and False Negatives further converted to A, P, R and F scores.
- The author now compares existing algorithms by the means of statistical techniques like t-test, z-test, F-test and Chi-square test.
- After the analysis it is found that the Random Forest with Time complexity of O (M*m*n* log n), where M is the Number of trees, n is the data size, m is the number of features, is the best technique for the classification in the Online Social Network (OSN).

- The author gives conclusive remarks about the future direction of this research, which is that Along with the development of new models and techniques, there is a need to compare the existing models to adopt more robust and secure frameworks. Further, the optimization of the implemented framework, the evolutionary, and other approximation algorithms can be used and scaled up.

5) Abusive language detection from social media comments using conventional machine learning and deep learning approaches

By
Muhammad Pervez Akhter, Zheng Jiangbin, Irfan Raza Naqvi

- This paper essentially deals with the rise of Hate speech and abusive language, with the rise of the Internet. Focusing on the Urdu language which is like Hindi.
- The model is trained and tested on two scripts, the first being Urdu and the second being Roman Urdu or rather Romanized Urdu.

- The author describes which are the people at receiving end of this cyberbullying and trolling, namely people from specific religions or nations.
- The author attempts to detect the abusive language in Urdu roman dialect with more than two thousand comments.
- The author gives a brief introduction to the Urdu alphabet and the basic method of writing.

- The authors discuss unsupervised learning methods, identifying what data can be "useful" or can act as training data to obtain output values.
- The author gives different approaches or Deep Learning model methods to detect comments by Machine Learning and Neural Networks like CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), BLST (Bidirectional long short-term memory), CLSTM (Convolutional long short-term memory).
- The authors also apply ML models to detect abusive language and compared the performance of these models with Deep Learning.
    - Naïve Bayes: Uses conditional probability and Bayes theorem.
    - K- Nearest Neighbours (K-NN)
    - Support Vector Machine (SVM)
    - Logistic Regression

- On performing comparisons on the above-mentioned models, the authors derive that CNN or Convolutional Neural Network outperforms the other algorithms on both datasets.
- The Deep Learning Model in general performed better than conventional Machine Learning Models.
- The author gives conclusive remarks about the future direction of this research, which is mainly the generation of more and more datasets, the area of interest for these datasets are various social media websites, for instance, the likes of Facebook and Twitter to detect hate speech. Further, another research direction is to explore hybrid models of DL and ML to detect abusive language