Forecasting using Machine Learning

Amit Kumar Agarwal, Manish Shrimali, Sukanya Saxena, Ankur Sirohi, Anmol Jain

Abstract: Weather has a numerous impact in our daily life and also gained researchers attention due to its massive effect to the human life. To protect ourselves from weather we need to predict the weather such as rainfall, humidity and temperature etc. The emerging machine learning techniques in the last few years coupled with large volume of weather observation data. With the help of previous data, we predict the weather by using machine learning technique. In this paper, we implement machine learning technique for weather forecasting. In present myriad data available around us. So, it is very important for us to analyse this data in order to fetch out some useful information and intent. This can be done by using machine learning and data mining. Machine learning is an internal part of artificial intelligence, which is used to design algorithms based on the relationships between data and data trends.

Index Term: Weather, rainfall, humidity, myriad, data mining

I. INTRODUCTION

Machine learning is an internal part of artificial intelligence. In ML, computer learns automatically from data & information using different computer algorithm. Computer don't need to explicitly programmed. These can be improved& change algorithm by themselves. According to the report, the size of machine learning market is1.03 Billion (USD)in 2016.This expected to grow 8.81 Billion (USD) by 2022, at a Compound Annual Growth Rate (CAGR) of 44.1%. There is a growing need of machine learning among companies for professionals and it is used all over the world. will help us to understand the ins & outs of machine learning.[8]

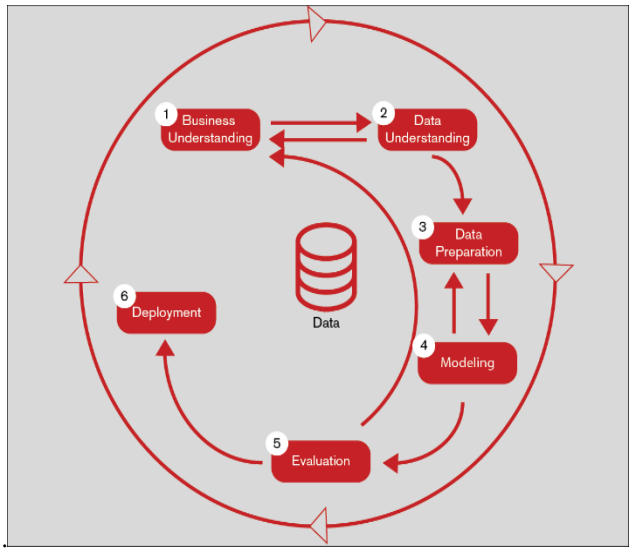

Fig 1. Steps involved in machine learning

Statistical models are refined and discovered in 1950`s. Machine learning research is conducted by simple algorithm. Bayesian methods are introduced for probabilistic inference in machine learning during 1960`s. Machine learning shift from knowledge driven to data driven approach. In 1990`s.Amazon launches its own "machine learning platform" and Microsoft created "distributed Machine learning Toolkit" in 2015. "Google artificial intelligence" algorithm beats a professional player at a Go (Chinese board game considered-as"world most complex board game")in 2016With the help of fig 1. We elaborate the steps involved in machine learning-

1. ***Gathering of Data***- Gathering of Data is very crucial step in quality and quantity point of view. It determines how good our predictive model is. The data is collected and converted into tabular form. This data is known as Training Data.
2. ***Data preparation***-. The next step is preparation of data. In this step, data is loaded into suitable place & ready to use in training of machine learning. Data is divided into 2 parts. The first part of the data is training data and other part of the data is testing data. These data sets are used for improving model`s performance.
3. ***Choosing a model***-After data After data preparation step, next step is choosing a model in which both data scientists & researchers have created over the years. Main is to choose correct model to get their job successful.
4. ***Training***-In this step, myriad training data used to predict the model`s ability & this process involves initialization of some random values say A & B for your model. With the help of these values, we predict the output of the model. After predicting the model. Our next step in this process is comparing the predicted value with model`s predictions. Then, adjust the values to match up with previously predicted values. 70-30 %, 60-40 %, 65-35% etc are some percentage of training and test data which we divide to verify our model.
5. ***Evaluation***-Evaluation is the next step after training step. This step involves testing of the model against the data. We basically compare the data which we used in the training step to verify our model with the new test data.
6. ***Parameter Tuning***- Once evaluation is done. Next step is parameter tuning, we basically used this step for further improvement in our training if possible. The learning rate is another parameter that tells how far away line is shifted in each & every step.

It is based on the information provided from the previous training step & plays crucial role in the accuracy of training model, & how long the training will take Differences can be seen depending on whether a model starts off training with values initialized to 0`s versus some non-O`s,

Revised Manuscript Received on December 22, 2018.
Amit Kumar Agarwal, ABES Engineering College, Ghaziabad.
Manish Shrimali, ABES Engineering College, Ghaziabad.
Sukanya Saxena ABES Engineering College, Ghaziabad

which then leads to the question of which distribution used. The adjustment of these parameters generally depends on the, model, dataset, & the training process.

7. ***Prediction***- Prediction is the final step in machine learning & used to answer questions. At this final point machine learning value is realized & you can finally use your model to predict the outcome of what you want earlier.

II. ARCHITECTURE OF MACHINE LEARNING

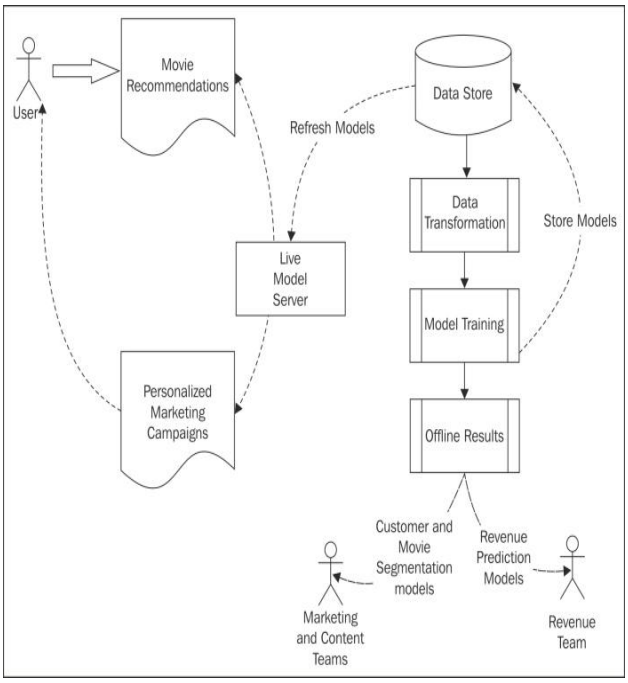

Fig 2. Architecture of Machine Learning

With the help of Fig 2. We elaborate the architecture of machine learning.

According to the diagram,

1.We need collect the data for users, their behaviour and content titles.

2.After collecting the data, transform this data into features.

3. Next step is to train our model by using training and testing data. Training also done with model selection phase.

4.Next step is to deploying the trained model to live-model serving system.

5.Model Feeding back is the results of Movie Stream website via targeting & recommendation pages

6. Model feeding back is the results into personalized marketing channels of Movie Stream's

7. Using the offline models to provide tools to Movie Stream's various teams to better understand user behaviour.

Machine learning is well-suited to gain competitive edge in digital business due to following reasons.

1. **Speed** to support faster compute calculations and decision making. Under the right sets of conditions, ML can be used to deliver valuable business insights more quickly and efficiently than many other analytics techniques because there's no need to program every possible scenario.
2. **Power** to process and analyse large volumes of data. ML can use higher volumes of data than those traditional techniques, and it has the potential to perform much more powerful analytics.
3. **Efficiency** to generate more models, more accurately, than traditional analytical and programming approaches. It offers efficiency in enabling models and insights to be generated without human resources or coding.
4. **Intelligence** through the ability to learn autonomously and uncover latent insights.

Different types of machine learning-

1. **Supervised learning** –Supervised learning is type of learning in which input and desired output is provide to the machine. Input & output data are labelled for classification purpose. Target variable is predicted from given set of predictors. Random forest, KHH, Decision Tree, Regression, Logistic Regression, boosting algorithm etc. are some algorithm used in supervised learning.

Types	Usage example in business
Neural Network	Predicting Financial Result & Fraud detection
Classification& regression	Spam filtering and fraud detection
Decision Tree	Risk Management, Threat management system

2. **Unsupervised learning**–In this type of learning, which is used to draw inferences that comes from datasets consist of input data without having labelled responses. We don`t have supervisor to guide us. Apriori algorithm, K-means, Adaptive Resonance Theory, Self-organising map (SOM Model) are some algorithm used in unsupervised learning.[9]

Types	Usage example in business
Cluster analysis	Financial transactions, streaming analytics in IOT
Pattern recognition	Spam detection, Biometrics, Identity management
Association rule learning	Bioinformatics, manufacturing & assembly

3. **Reinforcement learning**–Reinforcement learning is same as supervised learning that trains algorithms using a reward and punishment technique. A reinforcement learning algorithm learns machine by interacting with its environment. Supervisor will give reward at the time of correct action and punishment at every wrong action. Here, machine is trained. So that it will take its own decision. Reinforcement learning learns from their previous experience. To make accurate decision, it will capture the best possible knowledge. Example of Reinforcement learning- Markov Decision Process.

Depending upon problems, we have a more than one options to solve most of the problem. Before applying machine learning algorithm. We need to understand the problem. What type of data used in the problem.

Check the possibility whether another algorithm works same. Then, there is a possibility that second algorithm gives more accurate result. **Following are some algorithm in machine learning**.

1. Logistic Regression
2. KNN
3. Random Forest
4. Linear Regression
5. SVM
6. Decision Tree
7. Gradient Boosting algorithms
8. K-Means
9. Naive Bays
10. Dimensionality Reduction Algorithms.

Applications of Machine learning are-

1) **Social Media Services**
2) **Virtual Personal Assistants**
3) **Email Spam and Malware Filtering**
4) **Predictions while Commuting**
5) **Videos Surveillance**
6) **Online Customer Support**

As we all know forecasting around us. Machine or people have been forecasting weather pattern, sports outcomes, political events & many more. Using simple intuition, expert opinions, or using of previous[10]results to compare with the traditional statistical. Forecasting can be improved by adding newer machine learning techniques.

Issues of machine learning-

Machine learning advocate think the technology is a revolutionary method of fighting against security threats; even stopping attackers in their tracks *before* they've managed to attack a network. The more sceptical, or some would say, rational elements of the security community see it as a useful tool, but one that is not a solve-all for the industry's woes. [5]

Machine learning models can also be reverse-engineered. Researchers demonstrated in 2016 that they could replicate an Amazon ML model with near-perfect accuracy, simply by logging its responses to a few thousand queries. This has obvious implications for how attackers use training data and threat information simply by probing publicly available algorithms, and then create new attacks so that they appear normal to the ML-driven analysis.

III. LITERATURE SURVEY

–As we all know, rapid development in the fields of software, communication technologies& hardware. This will facilitate the emergence of Internet-connect sensory devices. This will provide data measurements& observations from the physical world. By 2020, it is observed that the total number of internet-connected devices being used will be in between 25 & 50 billion. It`s assumed that the internet-connected devices will grow and technologies become more matures compared with past. The volume of data being published will increase. The technology used in Internet-connected devices referred to as Internet of Things (IoT), continues to extend the current Internet by providing interactions& connectivity between the physical and cyber worlds [1]. In addition, as cheaper sensors and better connectivity expand the accessibility of the internet of things (IOT), the number of devices & pieces of equipment that can provide useful real-time information about the weather will likely expand dramatically. [4]

Data analysis plays a crucial role in the fields of machine learning& it became challenge for researchers and engineers in the fields of machine learning. "The real prerequisite for machine learning is data analysis, not math." means that data scientists spend its myriad time on data analysis only. According to traditional statement, data scientists "spend 80 percent of their valuable time on data preparation only." According to me this statement is essentially correct, a more precise statement stated that we will spend 80 percent of our time on fetching data, data reshaping, cleaning data, data aggregating, data exploring etc. These can be done by data visualization &exploratory data analysis. Ultimately, the importance of data analysis applies to both data science& ML too. Fact stated that if you want to create a ML model, you will have to spend myriad time just doing data analysis as a precursor for that particular process. Moreover, you will use data analysis to explore the model results. After that we have to apply an ML algorithm. Feature selection is an important process. It provides an effective way to solve these types of problems. Simply by removing redundant irrelevant data, which can improve accuracy of learning, reduce the computation time, & facilitate a better understanding for the learning model or data [2].

There are various techniques of predicting weather using Regression and variation of Functional Regression, in which datasets are used to perform the calculations and analysis. To Train the algorithms ¾ size of data is used and ¼ size of data is termed as Test set. For Example, if we want to predict weather of Austin Texas using these Machine Learning algorithms, we will use 6 Years of data to train the algorithms and 2 years of data as a Test dataset [3]

On the contrary to Weather Forecasting using Machine Learning Algorithms which is based primarily on simulation based on Physics and Differential Equations, Artificial Intelligence is also used for predicting weather: which includes models such as Neural Networks & Probabilistic model Bayesian Network, Vector Machines. Among these models Neural Network is comm. Only used algorithm due to its ability to capture non-linear dependencies of past weather trends & future weather conditions.

However, certain machine learning algorithms and Artificial Intelligence Models are computationally expensive, such as using Bayesian Network and machine learning algorithm in parallel.

The scope and availability of the weather-related data is massive. Thousands of weather satellites in space currently providing worthy data about winds, temperature, cloud patterns etc. These satellites produce modicum data. Hundreds of thousands of private & govt. weather stations on Earth, constantly gathering real-time data. Weather Underground Company claims to have to access 250,000+ personal weather stations. This will provide real-time information.

To conclude, Machine Learning and Artificial Intelligence has greatly changed the paradigm of Weather forecasting with high accuracy and productivity.

And within the next few years more advancement will be made using these technologies to accurately predict the weather to prevent disasters like hurricane, Tornados, and Thunderstorms.

For a bank to consider whether or not to offer someone a loan they go through a sequential list of questions to figure out whether it is safe to give loan. Those questions start from source of income, monthly income lies between some category, credit score etc.

A decision tree is largely used non-parametric effective ML modelling technique for classification& regression problems[6]
The problem of multi factorial decision presents in various domains. As everyone knows, a great majority of things results from multifactor not results from single factor in natural world. A choice made between alternative courses of action in a situation of uncertainty. You need to make a decision between red/blue paint for the sign that hangs outside of your store. When people make any decision, at first, they take the effect of every factor into account respectively, & then they make a synthetic decision of all factors again [4]. For example- you want to buy a car, we usually take budget, mileage, company, shape, colour etc. When we predict, we think about various factors; after analysing it, we make our weather prediction. [7]

Methodology –Beginning in July 2018, severe floods affected the south Indian states such as Kerala due to unusually high rainfall during the monsoon season. It was the worst flooding in Kerala in nearly a century. Over 483 people died, 14 are missing. At least a million people where evacuated mainly from Chengannur, Pandanad, Aranmula, Aluva, Chalked, Kuttanad and Pandalam. All 14 districts of the state where placed on red alert. According to the Kerala government, one-sixth of the total population of Kerala had been directly affected by the floods and related incidents. The Indian government had declared it a Level 3 Calamity, or "calamity of a severe nature". It is the worst flood in Kerala after the great flood of 99 that happened in 1924.

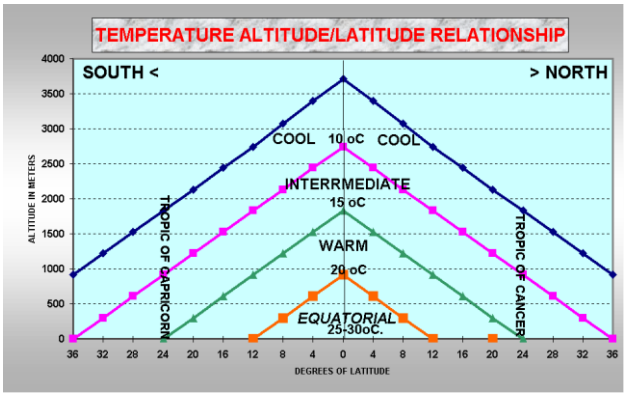

Fig 3. Describes the relationship between Latitude and temperature
These are caused due by low pressure and heavy rainfall. Kerala received heavy monsoon rainfall which is about 257% more than the usual rain falling in Kerala, on the mid evening of August 8 resulting in dams filling to capacity; in the first 24 hours of rainfall the state received 310 mm (12 in) of rain. Almost all dams have been opened since the water level has risen close to overflow level due to heavy rainfall, flooding local low-lying areas. For the first time in the state's history, 35 of its 54 dams have been opened.
Table 1: Describes Maximum and minimum Temperature

S No	Duration	Maximum Temperature	Minimum Temperature
1	June 2018	32 C	23 C
2	July 2018	32 C	23 C
3	August 2018	30 C	23 C
4	June 2017	33 C	23 C
5	July 2017	32 C	23 C
6	August 2017	32 C	24 C

From figure 1 we can see that Kerala is having **10.850516 latitude which is fix and temperature in Kerala is more than 30 C. probably, due to latent heat release during conversion of water vapour to rain up in the atmosphere.** There are two effects at surface, 1. Decrease in temperature of the atmosphere near surface due to rain. 2. Increase in water vapour in the atmosphere near surface. If the question is "why one feels warmer (at surface) immediately after the rain?" then I think, it is due to increased content of water vapour. The temperature felt by the body (difference between Body temperature and outside) increases since just after the rain the cooling mechanism of the body finds it difficult to generate sweat in moist environment. In this methodology, we generally considered temperature and latitude for checking how climate change. From above data, we analysis a correlation between temperature and latitude. How it is affecting the areas located nearer or further from the latitude. From above table, $RH=E/E_s * 100\%$.
E- Vapour Pressure E_s – Saturated Vapour Pressure.
Equations for Clausius-Clapeyrons

LN(Es/6.11) = (L/Rv) (1/273 - 1/T)
Es = Saturation vapor pressure
L = Latent heat of vaporization = 2.453 × 10^6 J/kg
Rv = Gas constant for moist air = 461 J/kg
T = Temperature in Kelvins

IV. CONCLUSION

In this research paper, after reviewing the above methodology with respect to the problem described in the Abstract section is exactly matched. We found that the relation between the temperature and latitude for a different region that effect the weather that causing heavy rainfall, flood and disturbance in the atmosphere.

REFERENCES

1. Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, Amit P. Sheth, https://doi.org/10.1016/j.dcan.2017.10.002 “Machine learning for internet of things data analysis: a survey”, Networks. Volume, August 2018, Pages 161-175
2. Scott C. James Yushan Zhang, Fearghal O'Donncha, “A machine learning framework to forecast wave conditions”, Coastal Engineering Volume 137, July 2018, Pages 1-10.
3. Jie Cai, Jiawei Luo, Shulin Wang, Sheng Yang, “Feature selection in machine learning: A new perspective”,Neurocomputing, Vol.300, pp.70-79, 2018.
4. WengianZhang, YomgviChen, “Mathematical Models of Multifactorial Decision and Weather Forecast”, IFAC Proceedings Volume 16, Issue 13, July 1983, Pages 265-269.
5. Joy E. Losee, Susan Joslyn, “The need to trust: Howthe forecasted weather influence forecast trust”, international Journal of Disaster Risk Reduction Volume 30, Part A, September 2018, Pages 95-104
6. Yuan Quan State Key Lab. of Intelligent Technol. & Syst., Tsinghua Univ., Beijing, China, “Research on weather forecast based on neural networks”, Procedia Technology Volume 4, 2012, Pages 311-318
7. Jitcha Shivang, S.S Sridhar, “WEATHER PREDICTION FOR INDIAN LOCATION USING MACHINE LEARNING”, International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 1945-1949 ISSN: 1314-3395 (on-line version).
8. R Usha Rani, T.K. Rama Krishna Rao, R. Kiran Kumar Reddy, “An Efficient Machine Learning Regression Model for Rainfall Prediction”, International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 23, April 2015.
9. Sanyam Gupta, “Weather Prediction Using Normal Equation Method and Linear regression Techniques”, International Journal of Computer Science and Information Technologies, ISSN-0975-9646, Vol. 7 (3), pp-1490-1493, 2016.
10. P. Kalaiyarasai, Mrs. A. Kalaiselvi, “Data Mining Techniques Using to Weather Prediction”, IJCST, Vol.6, 2018.