RESEARCH PAPER 2021

By Rishabh Chauhan

A2305219440  
Under guidance of

Dr Sunil Kumar

*Abstract*

**Big Data Analytics, Artificial Intelligence (AI), Internet of Things (IoT) and cloud computing all together has emerged with an ultimate goal of automating and changing human life by providing their services. These incredibly strong technologies have huge potential by working together, making human life simpler and advanced. To increase the popularity of any of these services, Quality of Service (Qos) metrics are needed to be defined clear. One of those quality metrics is packet loss or packet delivery, which is the main research idea of this paper. With advancement in Intelligent Network (IN) there exists a scope to predict packet loss, by analyzing the recorded network traffic and processing said data under certain machine learning algorithms to create a model to either predict packet loss or tell which variable is responsible for packet loss. This paper introduces all the said topics (Big Data Analysis, AI, IoT, Cloud computing) and conveys the relation between IoT, Big Data analytics etc and IN and QoS.**

*Introduction*

Big Data analytics is one of the most popular fields in current times. Big Data analytics as suggested in name is the analysis of big data i.e. to process data and uncover meaningful and hidden insights that can be used as a profitable asset by the company. By analyzing and visualizing the big data, organizations and stake holders can make correct decisions and resort to new techniques to meet the needs of their audience. To understand analytics, the concept of big data should also be clear. There are 5 Vs of big data known as veracity, volume, variety, velocity and value.

Intelligent network or IN is the term coined for a network that provide specific technical capabilities and services which are far more advanced than the primitive network. This network serves outside the basic network standard spectrum. The intelligence is placed in computer nodes giving network provider a way to upgrade, modify, develop and control services efficiently. Wireshark is free to use packet-analyzer software that can be easily installed on a system. This software records all the network traffic and gives a detailed report of a system network capabilities and flaws. Wireshark is used to generate the required data set for analysis.

Analysis of Big data is not an easy job for traditional computing power as in its name “big data” the data is quite large and requires a much power full compute power which can be accessed easily without any massive up-front investment by the help of Cloud Computing. Cloud Computing is nothing but providing compute power, storage, application etc on demand via the internet. This technology is the way to go in the current time being as it comes with a lot benefits that will be discussed later.

Artificial intelligence, considered being the hottest field in the technology industry. AI is the branch of computer science that deals with making machines to mimic or copy the working of human intelligence. XGBoost is an AI classifier that is used in the project to predict packet loss. The build of the prediction model is done on jupyter notebook and code is written in python programming language.

Internet of things is also one of the fields that are gaining popularity with time. It involves working with certain devices containing sensors which enable these devices to communicate with each other to transmit and gather data. Paper also talks about how QoS metrics are embedded into the architecture of IoT.

*Problem Statement*

**The problem statement is to somehow make a packet loss prediction which is not applicable generally but will be applicable and personalized to specific systems, hence predicting packet loss on any captured network data with wireshark.**

*Proposed Idea*

**The proposed idea for this research is to learn about Packet loss, which is one of the QoS metrics and making an AI model to predict packet loss with expected accuracy of more than 70 percent.**

*Big Data analytics*

Before diving into the sea of Big Data Analysis, it is important to know what Big Data means. Quite surprisingly, Big Data was considered to be a serious issue few years ago. In the early 2000s, CPU and Storage technologies were shocked by the terabytes of big data received due to skyrocketing of data volumes which resulted in data scalability crisis. With the help of Moore’s law, the computing power improved to have greater capacity, speed and intelligence while also being quiet affordable and hence enterprises had amazing data collection and analysis practices.

Big companies and enterprises are adopting various data analysis strategies to get insights out of data. This is quiet important because of economic recession lead to changes in various businesses. Using latest analytical tools and technologies, companies can study big data deeply to understand the hidden business insights.

So what actually is Big Data analysis, well in a short summary it is the use of advanced analytic techniques to mine hidden insights out of big data sets. So overall the two main ingredients are Big Data and Advanced Analytics.

Big Data is nothing but Data which is really large in volume and is still increasing exponentially overtime. This is data with huge size and such complexity that its storing and processing is borderline impossible for traditional data management tools. To understand big data with more depth, it can be described via five Vs of Big Data namely velocity, variety, volume, veracity, value.

* Velocity

Velocity refers to the speed or the rate at which the big data is collected or generated. With the advancement in hardware and network technologies, this speed is bound to increase which results in capturing of more data insights. This data flows in from sources like machines, networks, social media, mobile phones etc. For example, there are more than 3.5 billion Google searches per day.

* Variety

It is the different types of data that is captured. This can be classified into structured data, semi-structured data and unstructured data.

* Structured Data

It is the data that is organized neatly in formatted repositories, resulting in data being more effectively processed and analyzed. For example data in excel sheets or in tabular form.

* Semi-Structured Data

This data is basically semi-organized data. It does not conform to the formal structure of data. For example log files.

* Unstructured Data

Data that is not organized is called as unstructured data. It is the type of data that does not fit even closely to the traditional rows and columns structure of the excel sheets. The data obtained from a video, audio, pictures etc is an example of unstructured data.

* Volume

This is the main feature of big data as in its name “big”, it defines if the given data is big or not, so volume is the key characteristic of big data.

For example in 2016, the average mobile traffic was 6.2 Exabyte per month, Netflix has over 86 million members globally.

* Value

Data in its raw form is useless and unimportant unless it is converted into something valuable to mine hidden benefits from it. So it can be said that Value is one of the key feature of big data. Simply said, processing data to produce and bring out value is really important for big data architecture.

* Veracity

The data being inconsistent and uncertain is given the term veracity. This means that veracity is the trustworthiness, quality and accuracy of data.

Big Data Analysis Tools

As we have discussed about big data, it provides huge statistical samples which in turn enhances the advanced analytical results. Now all the computing tools for data mining and analysis are designed and manufactured to handle large data sets. There is a basic rule, “Larger the data set, more accurate the analysis results”. One of the tools is SQL which converts big data in search of the best suitable customer segment, churn profile, etc. Analytic tools like SQL can handle the big datasets and can also execute queries and parse tables in record time. These tools are also tolerant of raw data including non-standard and poor-quality data which is not bad as predictive analytics depend on lots of tiny details. Real-time Fraud detection relies on this non-standard data as the fraud indicators. There are many more data analytic tools like R studio, SQL Server, Oracle DB, IBM Watson studio, talend etc. on which data analysis can be achieved. There is no perfect tool for data analysis, the correct tool is the one which gets the task in hand complete, so a decent knowledge in all of these tools takes an average data scientist one step ahead of the crowd.

Advantages of Big Data Analytics

Big data analytics benefits customer relations, business intelligence, and much more analytic application. To stand out from the other companies, organizations are expected to have a unique approach to market their products. By harnessing the true power of big data, organizations can spot exactly what the customers need. Big data analytics observes patterns and uses the results to project their products to the customers who actually need it. Amazon provides one of the best personalized shopping experiences. There are endless advantages of big data analytics and a lot of them will be shown later in this paper on how the large data was collected and how it was analyzed.

*Wireshark*

Wireshark is a free open-source packet analyzer. It is used for troubleshooting network, analysis, and software and communication protocol development.

For this paper WireShark is used to record the network traffic on my system. It records the source IP, Destination IP, time for each packet to be sent, tcp sequence and acknowledgement number and much more information. All this information is recorded to learn all the networking terms for understanding advanced networking or **Intelligent Network (IN)**

* TCP analysis

TCP stands for Transmission Control Protocol which is a communications standard that allows computing devices and application programs to interchange messages over a network. It sends packets of data across the internet and makes sure of the successful delivery of the given data and messages over networks.

Next expected sequence number

The last-seen sequence number plus segment length. Set when there are no analysis flags and for zero window probes. This is initially zero and calculated based on the previous packet in the same TCP flow. Note that this may not be the same as the tcp.nxtseq protocol field.

Next expected acknowledgement number

The last-seen sequence number for segments. Set when there are no analysis flags and for zero window probes.

TCP discussions are supposed to be finished when they have both opening and shutting handshakes, freely of any information move. Anyway, its better recognizing total discussions with some information sent, and we are utilizing the accompanying piece esteems to assemble a channel esteem on the tcp.completeness field

1: SYN

2: SYN-ACK

4: ACK

8: DATA

16: FIN

32: RST

* Packet Loss

Sadly not all networks are awesome. This is particularly valid for This present reality, and that implies that occasionally packets sent by the network won't ever show up at the appropriate objective. Far and away more terrible, once in a while packets will be reordered by the network, or even copied by the network. In certain circumstances when packet loss happens there can be a critical exhibition debasement and accordingly it very well may be fascinating to us to attempt to limit the measure of packet loss happening.

For most organizations, packet loss is a regular conduct, for example this will occur if a Switch is getting more information than it can communicate.

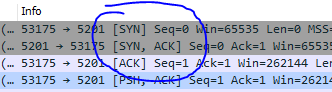
Now and then, imperfect equipment/programming essentially "neglects" packets.

In the event that the organization is designed effectively, there's very little that should be possible against packet loss as this is a fairly "planned" conduct.

* TCP 3-Way handshake

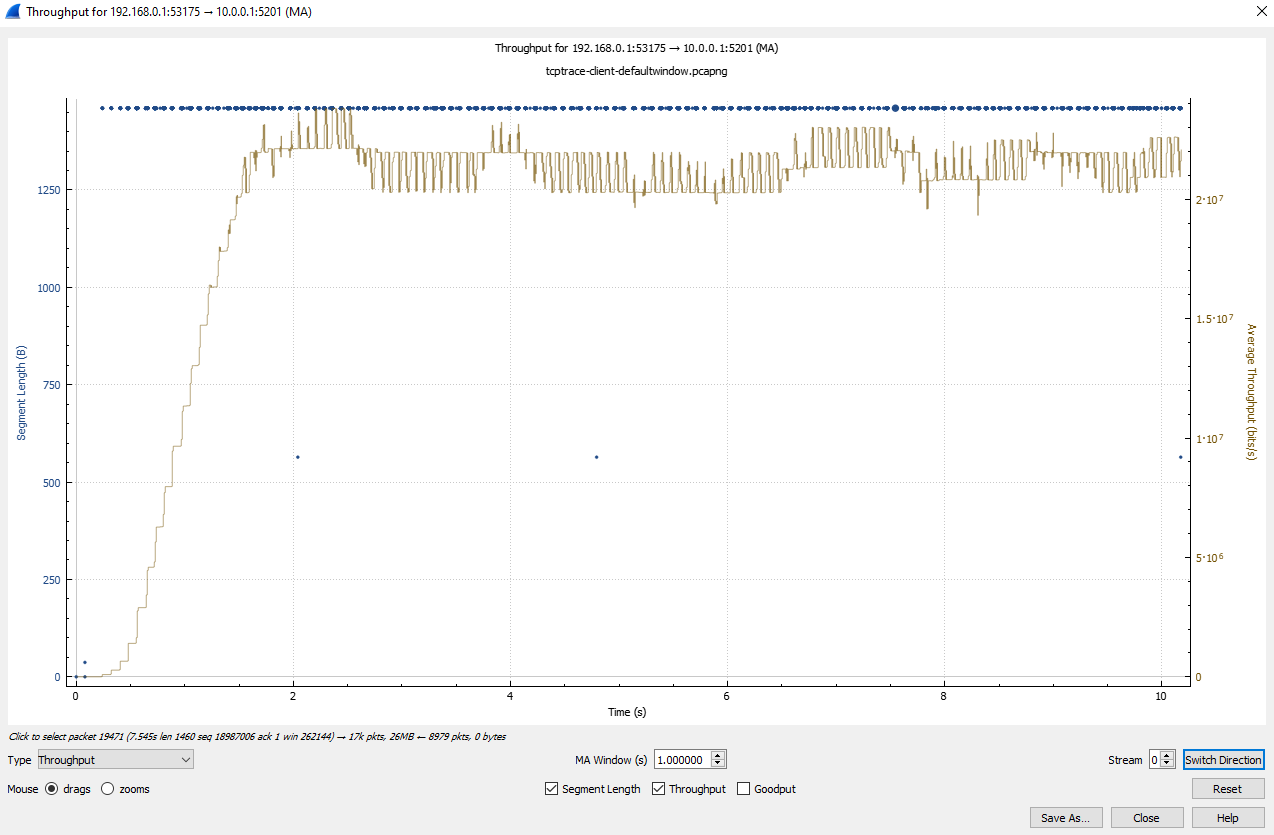
The cycle of communication between gadgets over the web occurs as indicated by the current TCP/IP suite model(stripped out variant of OSI reference model). The Application layer is a top heap of pile of TCP/IP model from where organization referred to application like internet browser on the customer side build up association with the worker. From the application layer, the data is moved to the vehicle layer where our point comes into picture. The two significant conventions of this layer are – TCP, UDP(User Datagram Protocol).

* Step 1 (SYN) : In the first step, client wants to establish a connection with server, so it sends a segment with SYN(Synchronize Sequence Number) which informs server that client is likely to start communication and with what sequence number it starts segments with
* Step 2 (SYN + ACK): Server responds to the client request with SYN-ACK signal bits set. Acknowledgement(ACK) signifies the response of segment it received and SYN signifies with what sequence number it is likely to start the segments with
* Step 3 (ACK) : In the final part client acknowledges the response of server and they both establish a reliable connection with which they will start the actual data transfer



* Throughput

Throughput measures how many packets arrive at their destinations successfully. For the most part, throughput capacity is measured in bits per second, but it can also be measured in data per second.



Wireshark captures packet on the system via communicating through Ethernet. A generalized speed test is commenced on the internet so that the upward (upload speed) packets and downward (download) packets are recorded. In this software the lost packets are shown with a black highlight and are of mainly two types- TCP Retransmission (fig. 2) and TCP Duplicate acknowledgement (fig. 1) (Dup Ack). This data is exported as a “.CSV” file and all the packets that are retransmitted or duplicated i.e. dropped packets are given the value ‘1’ while all the other values are given ‘0’ (fig.3) this is done to make the model understandable by the classifier and also to the user. Also some string values are also removed as they would cause problems making predictions.

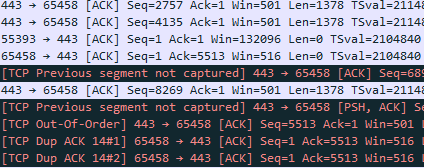


Fig.1 TCP duplicate Ack

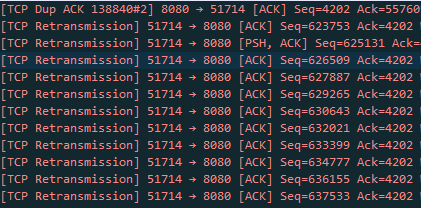


Fig.2 TCP Retransmission

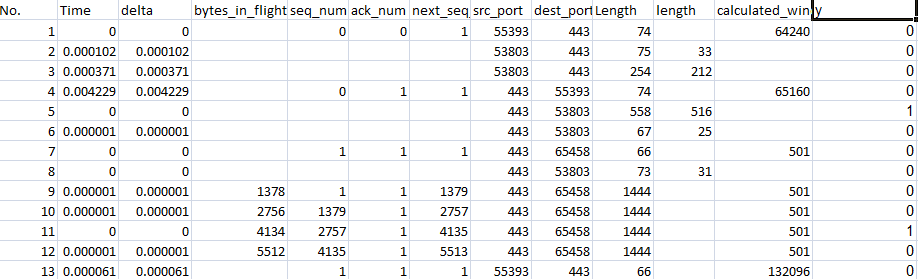


Fig.3 Processed Excel file

*Artificial Intelligence*

Artificial intelligence is one of many branches of computer science which deals with construction of smart machines achieving objectives requiring human brain. AI in itself is quite a vast field but machine learning (ML) and Deep Learning are creating the biggest impact in almost every company or industry. AI is based on the principle that how machine can copy or mimic the thinking and working of a human intelligence.

Artificial Intelligence is generally divided into two types- narrow AI and general AI

* Narrow AI

Nowadays, Narrow Ai fills up most of the Artificial Intelligence. This type of AI is designed to achieve one job at a time and to continue improving its execution. The objective is to find an automated solution or to just improve working software and make it much better.

* General AI

Also known as the ‘True Ai’ and considered to be the next step towards more complicated and advanced AI. In contrast with Narrow AI, this doesn’t focus on a single problem but to comprehend on a wide level. In simpler words this type of AI makes decisions not on training but on learning, this is how it mimics humans as it learns the way humans learns.

All this about AI is possible after understanding **Machine Learning** which is important for achieving Artificial intelligence. Machine learning is what is used to apply AI that allows software and system to advance depending upon their experience. There are 4 types of machine learning:

* Supervised Learning

Supervised Machine Learning can take what it has found out within side the beyond and practice that to new statistics the use of labeled examples to expect destiny styles and events. It learns via way of means of express example. Supervised Machine learning calls for that the set of rules’ feasible outputs are already regarded and that the statistics used to teach the set of rules is already labeled with accurate answers. Supervised gaining knowledge of is normally utilized in programs in which historic statistics predicts in all likelihood destiny events. Supervised learning is also divided into -Classification and Regression.

* Unsupervised Learning

This type of learning does not require any historical labels to be present in data and the system is not given any type of already known set of output or any interrelation between input and output. A transactional type of data is suitable for unsupervised learning. For example, identifying groups of people having common interests and characteristics. Unsupervised learning being so complicated and complex is used quite fewer times than supervised learning and only because of this reason, there is a groundbreaking future for advancement in unsupervised learning

* Semi-supervised Learning

As the name suggests, this learning falls somewhere between supervised and unsupervised learning. This is because some problems require a balance of both the learning. These are the cases when the data regarding the problem is given but for some reason it is incomplete, flawed, inaccurate or missing. This is when semi-supervised learning is brought into picture as it can access the given data and then use the unsupervised learning techniques to do the rest of the job. SSL reduces human bias. A labeled supervised learning AI is developed solely by a human and hence poses some risk of human error of improper or wrong labeling but with SSL and a lot of unlabelled data, the end result is much more precise and takes much less time and resources.

* Reinforced learning

Reinforcement learning or reinforced learning is a dynamic learning that trains a system based on reward and punishment. This learning algorithm learns by interacting with its surroundings. The system is rewarded for completing an objective or by performing correctly and is penalized for doing something wrong, hence learning like an actual human learns. Soon it seeks to earn or win and reduce casualties. Gaming, robotics, and navigation are some of the applications of reinforcement learning. Through trial and error, the algorithm determines which steps result in the highest rewards. The problem is known as a Markov Decision Process when it is repeated.

Advantages of Artificial Intelligence

As artificial intelligence applications are enormous and vast, this technology holds the power of revolutionizing the future. The advantages AI holds against all the other technologies is because of the following reasons-

1. Reduction in human error:

This is the biggest advantage that artificial intelligence possess because human error is one of the most common type of problem that exist at all times in a system causing some sort of casualty. With AI, the decision making process relies on the data that is previously captured or recorded by the machine or algorithm. So errors are reduced over the period of time and the system or program becomes more and more precise and accurate.

1. Saves human test-risks:

This is one of the reasons that can put to rest, all the negative and bad allegations towards AI being some machine taking over the world. Let’s take the example of one of the most dangerous accidents ever happened in history, The Chernobyl Nuclear power plant explosion. This accident happened due to heating overload in plant causing a lot of human casualties. If there were AI robots at that time controlling the fire and heat in first stage no disaster would have took place. AI can do all the risk involved activities without putting any human life in danger.

1. 24-7 Availibility:

Facts state that an average human being is likely to work for 4-6 hours a day. It is a basic human tendency to take some time out for themselves and dedicate this time to recreation activities, prepare for the next day at work and get plenty of holidays too. Unlike human beings, machines can be designed and utilised to work 24x7 using Artificial Intelligence. Machines neither require breaks, nor do they get bored while working.

1. Digital Assistant:

Digital assistants can be used by companies and organisations which will interact with users and this in turn will save human resources. Such assistants assist users while navigating websites and helping them look for whatever they need, via chatting mode. Some chatbots are designed in such a way that it is difficult to identify whether the user is interacting with a chatbot or a human being.

1. Faster Decisions:

AI and various other technologies can aid machines that make faster decisions and carry out functions quicker than a human. AI works by programming and delivers output in a faster and efficient way, unlike humans who are conditioned to consider many factors like emotional and practical before making any decision.

*Cloud Computing*

Cloud computing is the delivery of compute power, database storage, applications and other IT resources on demand, by using cloud services via the internet and with pay-as-you-go pricing.

Cloud can be defined as a computer located somewhere else to be accessed and utilized via the internet in some way. It can also be regarded as ‘web services’. It comprises server computers that are located in large data centers across the globe.

For example: Amazon Web Services (AWS) is a cloud service provided by computers owned by AWS to be utilized by people.

The computers provide various technological features and services such as building blocks, which can be used to assemble solutions that help a user to meet their business goals and technological requirements. Cloud computing allows organizations to utilize on-demand computing services and storage resources, rather than building, operating, and improving infrastructure on their own.

Before Cloud Computing

Cloud computing enables a hassle free use of compute power and storage, allowing users to pay only for the resources in need. Before cloud computing came into existence, people resorted to the traditional means of storage and compute power which was physical storage systems like storage servers, etc. Because of this the user has to pre-plan the capacity by guessing the theoretical maximum peaks. This is not the only problem as hardware solutions are physical, hence require space, staff, security and capital expenditure. All of this brings up the most important issue with traditional computing, which is significant up-front investment for buying, establishing and maintaining on-premises infrastructure. Cloud Computing on the other hand can get a business up and running with a significantly low up-front investment.

Utilizing Cloud Computing

Software is flexible and thus using infrastructures in terms of software offers a number of benefits. Software is capable of adapting to changes much more quickly, easily, and cost-effectively than hardware.

Software is capable of providing you exactly what you need with a few clicks in a few minutes. Cloud services provider like AWS don’t require you to plan your hardware needs ahead of time and then order, install, and set it up at your data center or undergo a long procurement cycle. This allows you to provision and terminate resources as necessary on AWS, without paying for hardware even when it is not in use. You can dispose of the resources whenever required to do so as they are temporary.

Powerful software like AWS allows you to be more agile and efficient in terms of managing change, testing, reliability and capacity planning.

The Three models of Cloud Computing

Cloud services can be a part of one of the three primary categories, which depends upon the degree of control and responsibility that one has over how the services are configured.

1. IaaS (Infrastructure as a Service) allows you to manage the server (physical or virtual) as well as the operating system (Windows or Linux). This means that the data center provider has no access to your server. The basic building blocks of Cloud IT are: networking features, compute, and data storage space.
2. PaaS (Platform as a Service) has someone else managing the underlying hardware and operating systems. This means that you can run applications without having to manage the underlying infrastructure (such as patching, updates, maintenance, hardware and operating systems). PaaS also allows developers to create customized applications with the help of a framework.
3. SaaS (Software as a Service) allows you to manage your files, however the data centers, servers, networks, storage, maintenance, patching, etc. is handled by the service provider. You are concerned about just the software and the way it has to be utilized. Examples of this are: Face book and Drop box.

The Three Cloud Deployment Models

1. All-In cloud : it is an application that is fully deployed in the cloud and all the parts of the application also run in the cloud. The cloud applications are either created in the cloud, or are migrated from another existing infrastructure. Such applications can be built on low-level infrastructure pieces (for example, networking, compute or storage) or high-level services can be used which provide an abstraction from the management, architecting, and scaling requirements of core infrastructure.
2. Hybrid : it is a way with which infrastructure and applications can be connected between cloud based resources and existing resources (the ones that are not located in the cloud). The most common way to deploy the hybrid method is between the cloud and existing on-premises infrastructure (also known as on-prem). This model is efficient in extending an organization’s infrastructure into the cloud by connecting the resources of the cloud into an internal system.
3. Private Cloud (On-premises) : private cloud is called so because here the cloud infrastructure is run from your own data center. It is useful for organizations who need to meet any compliance standards. In some cases, private cloud is the same as legacy IT infrastructure as application management and virtualization are used to increase resource utilization.

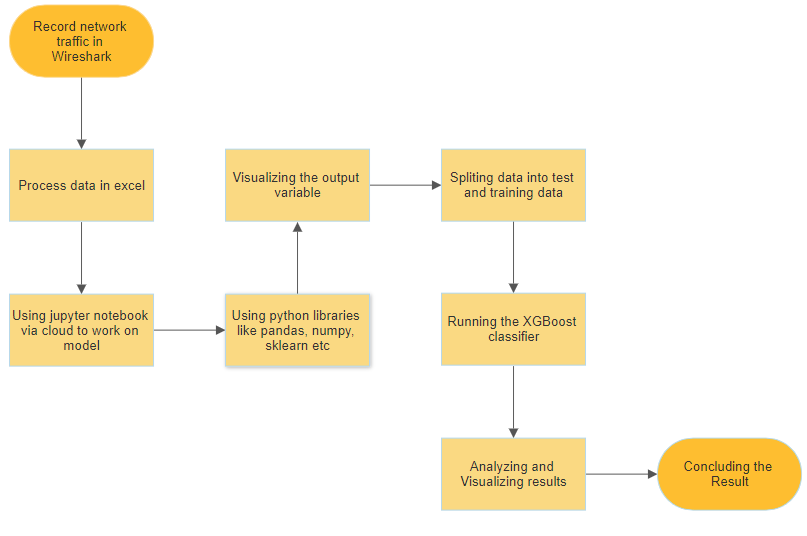
Advantages of Cloud Computing

1. Trade Capital Expense for Variable Expense: cloud computing allows you to pay for the resources when you use them, and for how much you use them. This minimizes the problem of investing heavily in data centers and servers without actually using them.

1. Benefits from massive economies of scale: cloud computing helps in minimizing the variable cost than what you achieve on your own. Data centers require physical hardware in terms of space, staff and security. On the other hand, when the customers are aggregated in the cloud, it results in a better economy.
2. Elimination of guessing the capacity needs: cloud computing helps to eliminate guessing on your own infrastructure needs. This means that one does not have to think about and guess how many resources would be required to accommodate maximum usage peaks, sufficient storage amount etc,. With cloud computing, you can scale up or down your resources with only a few minutes' notice.
3. Increase speed and agility: with the help of cloud computing, resources can be made available to developers within a few minutes only. This results in agility and efficiency of the organisation by proving global reach and rapid availability of new resources. You can also change your technology in minutes, whenever required. This also helps in safe experimentation of new technologies and ideas.
4. Strategic expenditure: cloud computing means that one does not have to spend money to run and maintain data centers. It lets you focus on your own customers and your core business  instead of the heavy lifting of racking, stacking and powering servers.
5. Ease of deployment (Go global in minutes): cloud computing helps in the easy deployment of your application in multiple regions of the world with just a few clicks. This ensures efficiency, minimal cost and better experience for the customers.

*Flow Chart*

The flowchart in the following figure shows the working procedure from recording data to concluding the result of the prediction model.

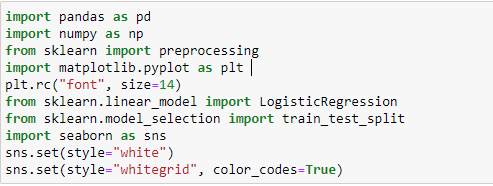


*Implementation*

The model for proposed idea is created on jupyter notebooks and the data set used for it and all other information is available on the <https://github.com/Rc17git/NTCC>

Libraries

There are a few libraries used for this model namely, Pandas which is a fast and powerful open source data analysis and manipulation tool. Numpy is also used which carries out all the scientific calculations along with XGBoost which is an implementation of a gradient boost decision tree developed for speed and performance.



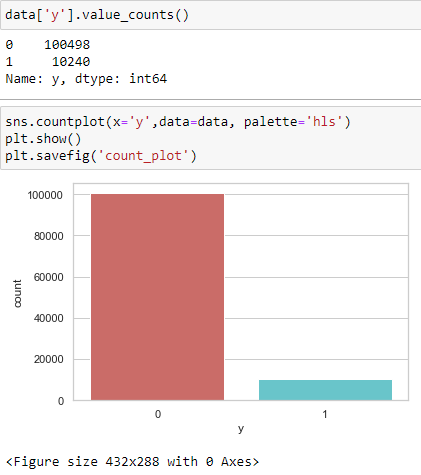
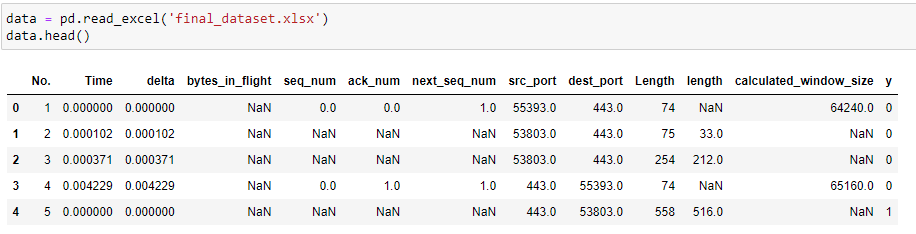
Dataset used

The data set used is manipulated and processed using excel and is captured with wire shark. The original data set contained 25 column which is reduced to get final data set of 110738 rows and 12 columns.

The main columns are seq\_num, ack\_num, next\_seq\_num which are indicators or response numbers which are generated when packets are sent or received. As the data set originally lacked throughput, jitter etc because of absence of variables to captured or calculate these values, these indicator numbers became the main variables for the model.

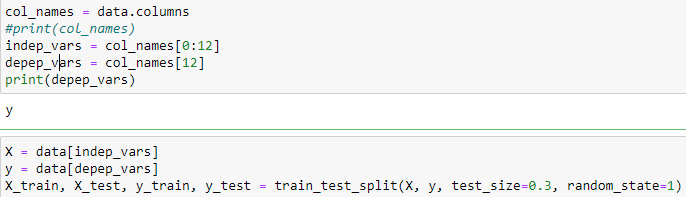
‘y’ field in the dataset has binary value 0 or 1, meaning ‘1’ for packet being dropped or lost and ‘0’ for successful packet delivery.

Also fields like src\_port and Dest\_port contributed to the accuracy of the model.



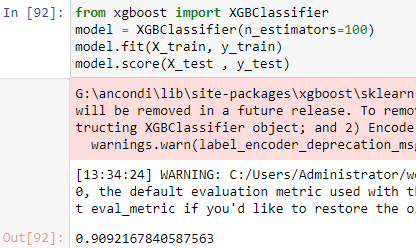
Visualization of no packet loss and packet loss

These types of classifiers work on the procedure of training data and test data. This means the model which is the XGBoost classifier is first trained with a training data and then after training of model it is performed on a test data. Splitting of data is done according to the “80 20 rule”, which means 80 percent of whole data is used for training while 20 percent is used for testing. In this case the data is split between 70 and 30 because packet loss is random and ratio of packet lost to packet delivered is pretty small. To split data the following code is used.

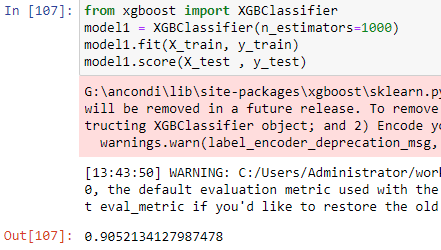


XGB classifier

XGBoost is an open-source software library that provides a regularizing gradient boosting framework for C++, Java, Python, R, Julia, Perl, and Scala. It works on Linux, Windows, and macOS. The project description aims to provide a "Scalable, Portable and Distributed Gradient Boosting Library".In this case we are using n\_estimators equal to 100 as well as 1000 just to increase the score. This classifier is best suited for the used dataset because it can handle infinite or NaN values which is nothing but a field having no value recorded for that packet or time. This classifier works on data set and gives an accuracy of approximately 91 percent.



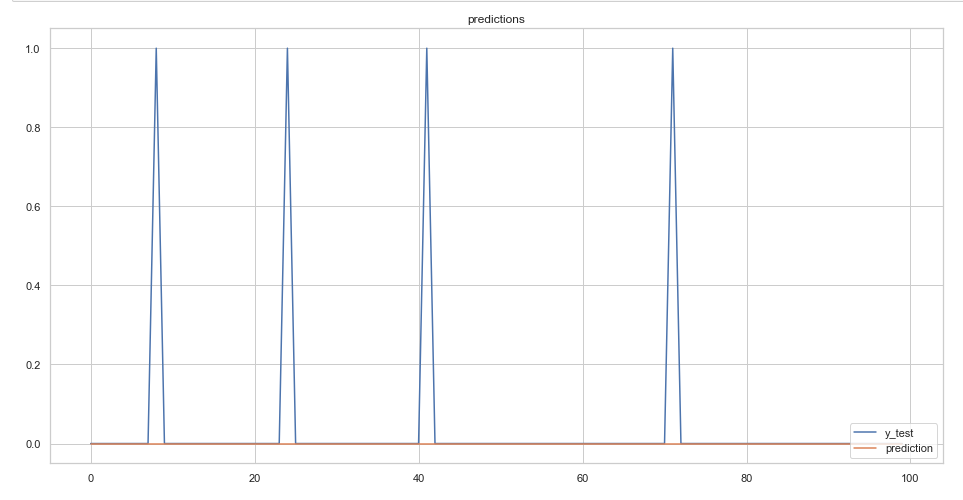
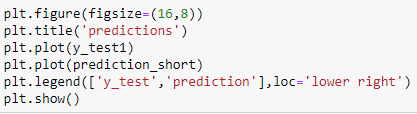
n\_estimators= 100



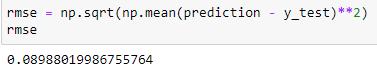
n\_estimators= 1000

Analysis and Visualization of result

**The accuracy of the model is 90 percent** and the actual vs. predicted graph is plotted using the following code



The graph above shows that the prediction overlaps the actual outcome and proves the system being accurate 90 percent of the time. It is plotted for 100 values actual outcome ( y\_test) and 100 values of prediction.



The model has a **root mean square error of 0.089** which is low and hence a good sign of a successful prediction.

The model is correct and can successfully predict packet loss for the data I recorded.

*Internet of Things (IoT)*

Internet of things or IoT, refers to the numerous physical devices that surround us and are connected to the internet. These devices are interconnected to each other and are communicating, sharing and collecting data. All these devices connect or communicate to each other using unique identifiers known as UIDs. “Things” in IoT can be anything ranging from a person with a heart monitoring implant which sends real-time data for analysis, a car with sensors to alert the driver in case of potential car accident or any other object that can be given a specific IP address and has capability to transfer data over the network. With advancement in technologies and time, a lot of industries and organizations are using IoT to work efficiently than ever before by understanding customer needs, better decision making and overall business growth.

Architecture of IOT

The IoT technologies are still evolving into its advanced form and most of the present or existing architectures are developed in WSN (wireless network sensors) perspective. All basic IoT models can be understood in three layer architecture namely, application layer, network layer, and perception layer. Internet of things generally constitutes intelligent devices, sensors, network connectivity and interconnectivity and last but not least a machine-human interface. All of these together can reform technology by establishing a new dimension in the world of information and technology. There are three components of IoT known as IoT of things, communication and computing.

* Things

“Things” are devices containing advanced chips and sensors and are capable of connecting to the internet. Smart devices, sensors etc are a few example of devices that can communicate to each other from anywhere in the world at any point of time.

* Communication

This component is designed to provide the communication in between the smart devices and also with the environment dependent upon the same communication protocols used.

* Computing

Computing or the compute power is required to analyze and work on the big data collected from the things. These computing capabilities can be acquired by cloud nodes, edge devices etc. Computing is the brain of IoT and it makes the correct decisions to get meaningful and profitable results out of IoT.

How QoS and IoT work together

Quality of service (QoS) provides IoT services by managing its resources and system peaks. With QoS the services providers are given a clear visibility of the services and performance that they can provide to the consumers. Using QoS a service level agreement can be enforced between IoT providers and IoT consumers. The QoS metrics are related and embedded in IoT components as follows-

* QoS of Communication

Communication networks in IOT deal with worldwide transportation of real-time data and applications. Such applications can either be delay sensitive or delay tolerant. In order to meet the requirements of a wide range of IoT applications, it is important to improve the quality of the network services by adding value to it. Some qualities of a good communication network are:

1. Jitter:

jitter refers to an undesirable effect caused by an unequal deviation in the time intervals of the travelling data packet. The delay becomes unequal instead of remaining constant. Some possible causes of jitter include improper queuing, configuration error or network congestion. Similarly, jitter can also lead to packet loss and network congestion. If the bandwidth provided by the network service provider is sufficient, it can reduce the jitter.

1. Bandwidth :

bandwidth, measured in mbps, refers to the amount of data that can pass through a network at a given time-period, where the actual network speed could be much slower. A good network is the one which has high available bandwidth.

1. Throughput (efficiency) :

It refers to the measure of a number of data packets sent or received via network. Thus, it can be defined as the actual bandwidth which is available to a network.  It is measured in bps. As network latency increases, the network throughput decreases.

1. Network connection time :

It refers to the server connection time, which is the time taken by the server to reply to a data request made by some other device. If the server’s request is not received in a given time, it refers to connection timeout.

* QoS of Things

Diverse set of devices such as those embedded with sensors, equipped with RFID technology, and having a smart network connection are provided by the Internet of Things. Some important parameters to ensure the quality of things, i.e., sensors which would add a new array to the IoT applications are:

1. Weight:

As weight is a physical parameter, sensors with low weight and cost are preferred over those with heavy weight. Weight is measured in grams.

1. Long term stability:

Stability means consistency. A stable sensor would produce consistent output over time-period. Component ageing, noise etc., can lead to decrease in stability, depending upon the nature of the sensors.

1. Response time:

It is the time taken by sensors to change its output state with the change in output parameters. Sensors should have a low response time to function efficiently.

1. Precision:

When some data or signal is measured repeatedly under similar conditions, the deviation obtained by the sensors should be highly correlated to imply high precision.

* QoS of computing

An important component of IoT is Cloud Computing which increases the efficiency of applications generating vast amounts of data. Some important parameters of QoS which can be useful for users to compare their service providers are:

1. Scalability:

It refers to producing maximum throughput in minimum response time. In order to incorporate more users, a highly scalable computation power is required, which can be done by adding resources on the system.

1. Dynamic availability:

In a normal operating condition, dynamic availability implies whether the system is accessible to use or not. Availability in computing terms can refer to hardware and software quality being optimal and operational.

1. Pricing:

This criterion also plays a role while selecting computer services. In the computing environment, cost serves 3 functions- network, storage and compute. Service providers who provide best QoS with minimum cost are preferred the most.

1. Security and privacy:

Physical components of IoT like Cloud node or Edge node are also vulnerable to threats and attacks, like illegal access of personal information of a user, disabling of network availability and addition of erroneous data etc. Security involves the confidentiality and the integrity of the data using certain security measures like crypto algorithms, key management, physical, data and network security support.

*Conclusion*

The paper concludes the introduction to the major topics Big Data Analytics, Artificial Intelligence, Cloud Computing, Internet of Things and their relations with intelligent networking and Quality of Service. These topics have been introduced well with all the history, future uses and advantages as well as other important terms briefly explained. These topics have been thoroughly studied and applied successfully by analyzing big network traffic data, processing of data, using AI and Machine Learning classifier “XGBoost” and hence making a model to predict Packet loss which is a QoS metric with an accuracy of 90 percent. The model is personalized to work on wireshark data. It is quite accurate but can be faulty as packet loss prediction is quite a difficult task and to work on such a limited data set, lead to some set back. Packet loss itself is really random and can be solely removed from the internet service provider itself. Packet loss occurs mainly due to network congestion or bad service quality from the service provider. Packet loss could have been predicted more accurately if variables like throughput, jitter etc could be recorded side by side by wireshark. So in conclusion the problem statement of the paper is solved and the proposed idea is met and completed.