

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com jmeit@outlook.com



# The Importance of Data Science in Technical Industry with Special Reference to Hardware/Software Industry

# Vijay Laxmi Kalyani, Pooja Jha, Vaishali Ameta

Vijay laxmi Kalyani, Assistant professor (EC department) Govt. women engineering college, Ajmer, Rajasthan , India vijaylaxmikalyani@yahoo.com

#### Pooja Jha

**Pooja Jha**, pursuing B.TECH (VI Sem) EEE department, Govt.women engineering college, Ajmer, Rajasthan, India pjha5221@gmail.com

#### Vaishali Ameta

Vaishali Ameta, pursuing B.TECH (VI sem), ECE department, Govt. women engineering college, Ajmer, Rajasthan, India vaishali.ameta@outlook.com

Abstract: The data is converted into information from long time ago. Information is the key for any business to excel and overcome to competitions. Earlier the data was not so large and it was computed and converted into information. As the industry grows and with the growing need of demand and supply in comparison to increasing population of the world. Companies are strategically focusing on the information so that they can make the best strategy to win. The new flavor of the old win named as "Data Science" which is dedicated to computing data and extract the valuable information for formulating strategies. In this paper first we discuss about data science and how it is beneficial for technical industry with special reference to hardware/software industry.

KEYWORDS: Data science, machine learning, AI, analytical tools, hardware /software parts, data analysis, result, application.

## I. INTRODUCTION

In today's world, data present everywhere. The data is converted into information from long time ago. Information is the key for any business to excel and overcome to competitions. Earlier the data was not so large and it was computed and converted into information. As the industry grows and with the growing need of demand and supply in comparison to increasing population of the world. Companies are strategically focusing on the information so that they can make the best strategy to win. The new flavor of the old win named as "Data Science" which is dedicated to computing data and extract the valuable information for formulating strategies. With the help of "Data Science" the technical industries can easily track our data for example where we go, when we go, what we buy, what we searching online, what we share, what we click, and who are our friends etc.. Scientist and researchers are continuously working on it and exploring the probabilities to get the full adoption of this new technology. In this paper first we discuss about data science and how it is beneficial for technical industries. We will also discussed that which type of technology

and tools are used for data science. In this paper we take an example of DELL Laptop with special reference to hardware/software industry. We collect the data from different vendors and then analysis will be done.

#### II. DATA SCIENCE

Data Science is a buzz word. It is a dominant and emerging branch of Science and technology for the modern era. This branch of engineering and science is the combination of data mining (processing, analyzing storing data), probability, statistics, algorithm, computer science, imagination and art of creating unimaginable. The idea of data science came from word 'data' (Big Data) that can be turned into various products, application and solution. It is used in almost every field like commerce, health, education etc. Data Science is still an early beginning of growth stage of its advancement. Data Science is prominent field which involves automated process to analyze huge volume of facts and data. This branch incorporates the set of field which include processing, modeling, statistics analysis, mathematics, advanced computing, social science, business, information and computer science design. It is difficult to manipulate the messy data which is available on large scale i.e. big data. Data is rising and is ever present. Now a days we collect data generated from android device, sensors, mobile phones, web records, detectors, gadgets, satellites and convert it into digital.

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com





Fig.1. Methodology used inData Science Source: https://goo.gl/images/tRVu8v[1]

# III. IMPORTANCE OF DATA SCIENCE IN TECHNICAL INDUSTRY

With emerging technology the growth rate of industries, companies and the individual in the era of source of information depends on how rapidly and satisfactorily utilize a large quantity of data into particular commodity or product. And approach towards the invention of new technology. Whether its for processing millions or trillions of data into an actionable form. Data science allows the technical industry to collect the data from almost everything. The data has always been around us. The technical industry have the ability to collect more data at a faster rate and to do something with it.

With the fast improvements and advancements in technology with the help of hardware/software and computing resources etc., the technical industry can be able to gather ample data from almost everything for example: with the help of smart phone the companies traces of your movements and locations, when you searching online then your online search behavior is stored, even which serial, movie, channel etc. you like and how you use television can be captured, when you shopping from superstore/shopping malls what item(s) you purchase and how to purchase everything is recorded, how and where you walk in the city everything is monitored by CCTV and so on. The technical industry can be captured our data from everywhere and all the time without us noticing it.

Therefore the technical industry is pulling in massive amounts of data from users on mobile apps and websites, tracking where they go, when they go, what they buy, what they share, what they click, and who their friends are. They are in the perfect position to use this data to "predict the winners" [3].

With the help of technology like Machine Learning, Artificial Intelligence (AI), Data Mining and allied fields help in providing systematic and scientific methods to provide useful

insights, predictions, recognitions and analytical solutions from this gigantic data. Hence we can combine two ideas together and say "Data Science" is important.

1. Data Mining: It is the method of sorting via large volume of dataset to recognize, identify and develop the relation to solve puzzling through the analysis of big data.



Fig 2. Data Mining Source: https://goo.gl/images/ookEm5[2]

It has various parameters:

- Association Rate
- Database
- Algorithm
- Classification
- Predictive Analysis
- Clustering

The concept of Data Mining is growing popularly in every organization present.

2. Machine learning is a subset of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data. The process of machine learning is similar to that of data mining [4]

It is a big part of a "Data Science. It is a set of algorithms that train on a data set to make predictions or take actions in order to optimize some systems. It uses an algorithms to learn from data sets. Algorithms are essentially a series of steps that lead to the completion of a task. With the help of data and algorithms, machine learning technologies make intelligent predictions or perform actions.



Fig 3. Machine learning

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com



3. Artificial intelligence (AI): AI is a building technology that behaves like a human. In our smart and connected world, machines are increasing learning to sense, learn, reason, act and adapt to the world. AI is the programming that learns by events and reprogramming the devices according to the last seen pattern and not by already programmed instructions it is a kind of learning by experience, it advances time by time and event by event and, recognizing the pattern of the events etc. machine will improvise and work for human beings, but when they follow similar pattern of doing monotonous things and while continuous upgradation in the learning pattern of AI could be threat for human beings it is somewhat shown in science fiction movies. AI is the continuous learning process in which machine learns the pattern of the events and upgrade itself for the future events [5].



Fig 4. Artificial Intelligence Source:https://www.edx.org/course/artificial-intelligence-aicolumbiax-csmm-101x-0#![6]

# IV. NEED OF DATA SCIENCE TOOLS IN TECHNICAL INDUSTRY

Profitable business is always a driver of technology, and data science is no different. Big data has the potential to be mashed into data sets to understand customers much better than relying on those unscientific experts who make a best guess on hunches and past experience. Instead of guessing at what to recommend to a customer, businesses can use data sets that tell them exactly what a customer wants depending on the season, weather, past purchases, geographic location, and life events [7].

Netflix and Amazon are well known example for their powerful recommendation engines that use not only what people want to buy, but also what they look at. Netflix had been growing quickly since its inception in 1998 and it had a fresh approach to an established industry. It had a strategic competitive advantage in the form of reams of digital data at its fingertips and a hunger to bring to bear the best data science techniques available. The clever people at Netflix developed an algorithm into which they could input the film views and reviews of all of their customers. The algorithm would then make a good guess

at what additional films any particular user was likely to enjoy. Data Science was putting money in the bank for Netflix [8].

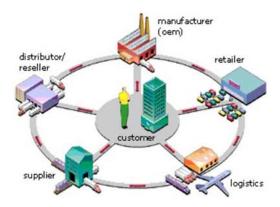


Fig 5. Business Cycle

Big Business, it turns out, really needs Big Data. And because of that, the tech industry is needed to harness the power of data science into something usable. Someone must create the apps and systems and algorithms that power these data-driven customer targeting engines [9].

# IV.I. Tools used in data science

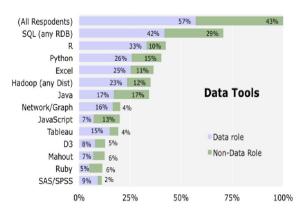
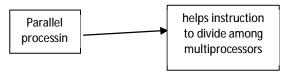


Fig 6. Data Science Tools Source: http://goo.gl/images/KcxZOJ[10]

1. R - R is most popular software package which is used to calculate math and visualization of results after collecting data. It is statistical modeling language which is freely available to users. It can undertake functionalized task like text analysis, speech analysis and genomic science. Now it has been upgraded for handling big data and parallel processing techniques. It includes-

 Parallel - This feature helps R take advantage of parallel processing.



Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com



- Snow This helps dividing calculation on a cluster of computers. Which is helpful in artificial learning.
- Rhadoop- This allows users an easy interface with hadoop from R.Map reduce is also a function of dividing the computing problems among different clusters and then recombine for getting a single answer.

Mostly R is used in companies like health care, finance marketing, pharmaceutical, bing, Facebook, bank of America.

2. JAVA and JAVA Virtual Machine (JVM) - JAVA is a variation of C++ and can be compiled once and run anywhere JAVA is being used in discovery, interpretation and communication of meaningful patterns in data. Mostly custom, analytical tools seeks this. Hadoop is also written in JAVA. JVM reduces the development time. It includes-

Scala - It is popularly used in data mining because it allows developers to use C++ as well as functional programming. Used by - The Guardian, LinkedIn, Foursquare, Novell, Siemens, Twitter, and the SPARK data mining environment at the UC Berkeley AMP Lab.

Clojure - It is jargon of artificial intelligence language LISP. It is designed for processing concurrent process. This can integrate with Hadoop and SQL.Used by - Netflix, Zendesk, Citibank, WalMart Labs and Spotify.

- 3. Python It supports data mining, statistical modeling and visualization. Used by Developers from Spotify, OKCupid and Evernote.
- 4. EXCEL -It has the ability to achieve a lot of sophisticated analysis and is easy to use. It is widely used in analysis, clustering, predictive modeling and optimization.
- 5. SAS (STATISTICAL ANALYSIS SYSTEM) SAS is a tough string that's popular for business intelligence analysis and unstructured datasets of large data. It is mainly used for advanced analytics, data management, and social media analytics. It is backward compatible hence easily handle historical datasets.
- 6. IBM: SPSS 6. IBM: SPSS MODELER AND SPSS ANALYTICS Its deal in all domains of big data modeling ie. loading, preparing, and predictive modeling, by using statistical or machine learning techniques.
- 7. SQL (Structured query language) and NOSQL -SQL is designed to collect more data and organize it neatly. But from survey it was found that more than 90% of data abide of articles, photos, videos, social media posts, machine-to-machine communication, product inventory, and technical documents.

Hence NOSQL is being introduced. Which with the help of Hadoop organize data .MONGODB is open source of NOSQL.

8.MYSQL - It is popularly used in data management of web applications like WordPress, Drupal, Facebook, Twitter, and YouTube.

9. Hadoop - It have the means to store and tackle the analytics for truly massive datasets. It is a free available software tool specially designed to help data scientists cope up with the unmanageable big data. It brings the analytics to the data by replacing the need to extract data from the storage devices altogether. It has increasingly become the industry standard for file system computing projects involving big data. Used by - Facebook, Yahoo, and The New York Times.



Fig 7. Data Science

Source: https://goo.gl/images/ihKITW[11]

- 10. A/B TESTING- Big data analysis involves finding complement in unexpected places in users data. Basically this technique works on comparison of websites against one another. Terms Used in A/B TESTING -
- Multivariate testing It allows to test more than two different versions of page against one another.
- Multi-platform functionality It ensures that our design or layout is user friendly and easy to use as well as performs well on every device.

Just like A\B testing, MaxymyserOptimizely and Adobe Target are also used for comparison purpose but are less effective than A\B testing so not used generally.

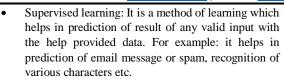
# 11. APACHE MAHOUT

Apache mahout is reliable, scalable, economic friendly. This branch is related to data mining and uses various techniques such as statistic, probability, recognition of shapes, patterns in computer vision. It is a software which is based on machine learning and various approaches are being used to solve the mysterious problems. In this argument I will focus more on supervised and unsupervised learning because the software mahout works mainly on these two types of learning.

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com



 Unsupervised: It is a methodology used for clustering, don't form prediction analysis.

# V. DATA SCIENCE IN TECHNOLOGY WITH SPECIAL REFERENCE TO HARDWARE/SOFTWARE INDUSTRY

To understand the Data Science. We take an example of dell laptop. We collect the data of 50 customers from various vendors those who are purchased dell laptop with different price ranges and different configurations. The price ranges of Dell laptop are varies from Rs.25000 to 60000 and processors are i3, i5 and i7 with different configuration like RAM, storage capacity etc. shown in table V.I,V.II,V.III.

••	Lapt op Dell Seri es	price range Upto Rs.250	price range Rs.25001 Rs.3500	price range Rs.35001 - Rs.45000	price range Rs.45001 - Rs.60000	Process or core i3	Proces sor core i5	Proces sor core i7	system memory- 4GB	system memory- 2GB	system memor y 8GB	system memor y- above 8GB		storag e capacit g 251GB- 500GB	ge	
1		1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
2		0	1	0	0	1	0	0	1	0	0	0	0	1	0	4
3		0	1	0	0	1	0	0	1	0	0	0	0	0	1	4
4		0	1	0	0	1	0	0	0	1	0	0	0	1	0	4
5		0	0	1	0	0	1	0	1	0	0	0	0	0	1	4
6		0	0	1	0	1	0	0	1	0	0	0	0	1	0	4
7		0	0	1	0	0	1	0	1	0	0	0	0	1	0	4
8		0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
9		1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
10		0	1	0	0	1	0	0	1	0	0	0	0	1	0	4
11		0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
12		1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
13		0	1	0	0	1	0	0	1	0	0	0	0	1	0	4
14		0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
15		1	0	0	0	1	0	0	1	0	0	0	0	0	1	4

Table V.I. Collected Data from various vendors

16	0	1	0	0	1	0	0	1	0	0	0	0	0	0	3
17	0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
18	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
19	0	1	0	0	1	0	0	1	0	0	0	0	0	0	3
20	0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
21	0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
22	0	0	0	1	0	0	1	0	0	1	0	0	0	1	4
23	0	0	0	1	0	1	0	0	0	1	0	0	0	1	4
24	0	0	0	1	0	1	0	1	0	0	0	0	1	0	4
25	0	0	0	1	0	0	1	0	0	1	0	0	0	1	4
26	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
27	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
28	0	1	0	0	1	0	0	1	0	0	0	0	0	0	3
29	0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
30	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
31	0	1	0	0	1	0	0	1	0	0	0	0	0	0	3
32	0	0	1	0	0	0	1	0	0	1	0	0	1	0	4
33	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4

Table V.II. Collected Data from various vendors

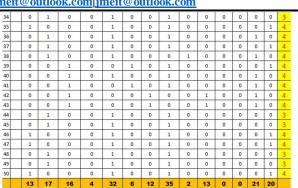


Table V.III. Collected Data from various vendors

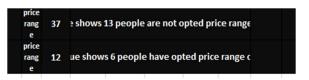


Table V.IV. Result from Collected Data

# V. RESULT

In table V.IV. We found that 13 people are interested to purchase the Dell Laptop upto range of Rs. 25000, 17 people opted the price range from 25001to 35000, 16 people opted the price range from Rs. 35001 to 45000, 4 people opted the range from Rs. 45001 to 65000. Therefore we conclude that the most demand of dell laptop in market i.e. varies the price range between Rs. 25001 to 35000 and customers interested to buy dell laptop with i3 and i5 processor.

The results shows that most people preferred the laptop of price range of Rs.25001-35000 & the most preferable processor, system memory & storage and capacity of device amongst the customer according to our research is i3, i5 with system memory 4GB and 8GB and storage capacity from 251GB-500GB and 5001GB-1TB.

# VI. APPLICATION OF DATA SCIENCE

It has wide application from business to medicine.

- Used in finance, govt. norms and policies, marketing analysis and use in every possible industries.
- Search engine specially Google process
- Digital marketing
- Used as recommender system for promotion of brands and products. For example; Amazon, Google, linked in, Netflix.
- Helps in recognition of image. For example; Facebook and Whatsapp.
- Urbanliving

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com



Fig 8. Emerging trends in data science Source: http://encrypted-tbn3.gstatic.com/images[12]

- Helps in fighting with crime via data
- For price comparison
- Predict flight delay
- Helps to solve fraudulent and in risk detection.

### VII. PROS AND CONS OF DATA SCIENCE

#### Pros -

- Large volume of data collected contribute in wide level for various causes.
- It uses various field like programming, machine learning, statistic, computation, probability. For which claim is to predict to grow and cultivate rapidly.

#### Cons -

- A large allocated amount of time is spent on analyzing and learning the data which is not existing.
- 2. Less probability of scope for longer investigation researches.

# VII. FUTURE SCOPE

Here we are discussing about how helpful is data science and has wide range of applications in every field.

Evolution - Within sixty or hundred years we can say that new set of new sets of creatures are going to emerge. There organisms will be very difficult from the present ones. They will be completely artificial in the way they will be designed by the technology. As the data rising day by day it shows that coming generation will be based on artificial intelligence. This will create an enormous change in the field of coming evolution.

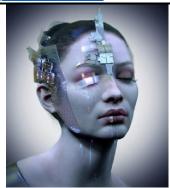


Fig 9. Future of Data Science Source: http://24.media.tumblr.com [13]

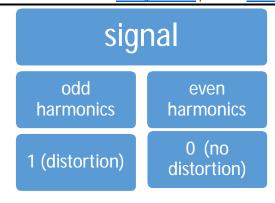
- Molecular Randomness This will help in the prediction of deformation in atomic and molecular structure. In simple words we can say that the understanding of physical as well as molecular structures and material properties along with the deformities will be easy.
- Product Development In this emerging trend of Data Science and Big Data is helping the engineering driven techie to develop various product by utilizing cloud based on Hadoop, SAS, machine learning and data analytics. This will be useful for startups.
- Sensex As the data is generated every second it grows. Every company's people take decision for investment on shares. By using the collected data by big data the probability in share market and banking sectors increases. Hence profit can be increased.
- Power harmonics As we know that removal of high order harmonics we use various kinds of filters but it is difficult to remove low order harmonics by using these filters. So for the removal of low order harmonics we can use data science and save energy. Thus we gate distorted free power signal. Which will work on predictive analysis and machine learning. It can be done by an algorithm, which will feed in the system software and by using this we will be able to find harmonics of low order which filter cannot remove.

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com





# Algorithm

- 1. Start
- 2. Declare variable H (Integer), B (Integer).
- 3. If (H%2==0)
  - a. B=0
- Else if (H>5)
- B=0

i. Else

ii. B=1

4. End

According to this when system displays output 1 which means signal is distorted (harmonics). When it displays output 0 it means signal is distortion free. When output 1 is generated that particular pace of an electrical signal get detected and pure signal will be sent for the transmission. So by using predictive analysis pure electrical signal and distorted signal can easily be identified.

Weather Forecasting - Data Analytics can be helpful in weather forecasting. Researchers uses the collected data and utilizes it for describing the science of atmosphere and to create computational mathematical models. It will help in prediction of floods, drought, hurricans and save millions of life and property loss.

#### IX. CONCLUSION

To conclude this it can be said that in this emerging field of engineering science and technology. Data science is the art of using the data effectively. It has immense probability to bring a change in coming era. Build the future to smartest tech. Its scope are huge and beyond our imagination. We suggest that the world explores its full potential and utilize it for the humanity and all over the development of nation.

Through this idea we have emphasized more in hardware & software domain so that they will be able to earn profitable to manage the product cost & features. Here we have successfully collected data from different vendors, the leading and emerging

brand which hold highest position in the global market. Through this research we concluded that price range between Rs.25001-35000 holds peak value. More numbers of customers are attracted towards this price range. Second best preference is hold by the customers who can afford to the price range of Rs.35001-45000. The laptop range of Rs.25000 & 45000-60000 are the least priority of customers. These are the market analysis of software & hardware products will help the company to manufacture their product according to the demand of the consumer. We have analyzed the data this will help the company to desgn their products whether it is hardware or software parts like Processor, RAM, ROM, Hard disk and many other parts. This data will not only help the specified company but other company also get benefited from this. Our idea will create revolution in product comparison, marketing analysis .Through this we successfully be able to predict the probability of customer choice towards particular product. According to the analysis of above data we concluded that company should design a device with these characteristics like processor i3 with system memory 4GB,8GB and storage capacity to 251GB-500GB in minimum cost price i.e. Rs. 25001-35000.

#### REFERENCES

- [1.] https://goo.gl/images/tRVu8v
- [2.] https://goo.gl/images/ookEm5
- [3.] https://www.import.io/post/why-data-science-is-the-fastest-growing-industry-in-tech-right-now/
- [4.] whatis.techtarget.com/definition/machine-learning
- [5.] Pawan Kalyani, "IoT Internet of Things, Artificial Intelligence and Nano Technology a Perfect Future Blend", Journal of Management Engineering and Information Technology ,Volume -2, Issue- 2, Apr. 2015, ISSN: 2394 – 8124
- [6.] https://www.edx.org/course/artificial-intelligence-ai-columbiax-csmm-101x-0#!
- [7.] https://www.import.io/post/why-data-science-is-the-fastest-growing-industry-in-tech-right-now/
- [8.] http://dsianalytics.com/data-science-privacy-and-netflix/
- [9.] https://www.import.io/post/why-data-science-is-the-fastest-growing-industry-in-tech-right-now/
- [10.] http://goo.gl/images/KcxZOJ
- [11.] https://goo.gl/images/ihKlTW
- [12.] http://encrypted-tbn3.gstatic.com/images
- [13.] http://24.media.tumblr.com

Volume -4, Issue- 2, Apr. 2017, ISSN: 2394 - 8124

Impact Factor: 4.564 (2015)

Website: www.jmeit.com | E-mail: editorjmeit@outlook.com|jmeit@outlook.com



# **AUTHOR'S DETAILS**

Vijay Laxmi Kalyani, H.O.D from ECE Department Govt. Women Engineering College, Ajmer. IAENG membership, attended and organized various seminars and conferences, STC and FDP'S etc. and published many papers in conferences and Journals.

**Pooja Jha** , presently pursuing B.TECH VI Sem. EEE branch from Govt.women Engineering College, Ajmer.

**Vaishali Ameta**, presently pursuing B.TECH VI Sem., ECE branch from govt. women engineering college, Ajmer

..