Efficient Clinical Decision Support System
Dr.P.Chitra\textsuperscript{1}, K.H.A iswarya\textsuperscript{2}, D.As wini\textsuperscript{3}, K.G.Geetha\textsuperscript{4}
Head of Department\textsuperscript{1}, Student\textsuperscript{2,3,4}
Department of Information Technology
Prathyusha Engineering College, Chennai, India

Abstract:
Computerized clinical decision support systems are information technology-based systems designed to improve clinical decision-making. CDSS a consistent challenge to physicians and is the area of research in which trillions of amounts are being spent by all countries. The Institute of Medicine proposes several recommendations to increase the quality of health care. Data mining algorithms have been applied on the voluminous health records to aid decision making process. In this system we make use of clustering technique (K-means), which provides clustering of health condition, patient health conditions are analyzed and exact disease is diagnosed using Big Data analytics which enhance both of productivity and efficiency of overall health services, success factors and barriers to successful CDSS implementation.

1. INTRODUCTION:
A clinical decision support system (CDSS) can be used by: health care professionals or physicians, in order to prepare a diagnosis and to reinterpret the already specified diagnosis and improve the final result. Healthcare businesses have witnessed a rapid growth by employing Information Technology in the automation of many of their processes such as transaction, saving inventory and maintaining records, which results in eliminating routine and repetitive processes. Healthcare can be defined generally as the management and the treatment of sickness with the aim of providing effective services to the patient. The amount of data produced in the healthcare business is huge; consequently, there is always a chance of incorrect diagnosis that may lead to drug reactions and allergies for a patient. Therefore, Information Technology is, used widely to capture and transfer information. Such system that helps the healthcare professionals in the problem solving process throughout the medical diagnosis is the Decision Support System (DSS) in order to confirm the correct diagnosis and reduce medical errors.

2. EXISTING SYSTEM
A clinical decision support system (CDSS) can be used by health care professionals or physicians, in order to prepare a diagnosis and to reinterpret the already specified diagnosis and improve the final result. The use of Clinical decision support systems has increased recently and showed an improvement in Productivity and performance, reduce medication errors and increase the quality of services in hospitals. Clinical data are huge and complex because of that analyzing them is not easy.

DISADVANTAGES
\begin{itemize}
  \item Complexity of the system as huge amount of knowledge base need to be extracted.
  \item Systems were not user friendly
\end{itemize}

3. PROPOSED SYSTEM
Decision Support System (DSS) can be defined as an examine a huge data and present the result of that examination and analysis to help users in making more effective and reliable medical decisions. This system is easy for the doctors to use, as it does not require command line inputs. They can view the result and appropriate provide treatment.

ADVANTAGES
\begin{itemize}
  \item Opportunity to save time by reducing decision time.
  \item Systems are user-friendly.
  \item Efficient and Proper Result.
\end{itemize}

4. SYSTEM DESIGN:
ARCHITECTURE DIAGRAM:
5. TECHNOLOGIES USED:

JAVA:
✓ Java is programmer’s language
✓ Java is cohesive and consistent
✓ Except for those constraint imposed by the Internet environment. Java gives the programmer, full control. Finally Java is to Internet Programming where c was to System Programming. Java has had a profound effect on the Internet. This is because; Java expands the Universe of objects that can move about freely in Cyberspace. In a network, two categories of objects are transmitted between the server and the personal computer. They are passive information and Dynamic active programs. But Java addresses these concerns and by doing so, has opened the door to an exciting new form of program called the Applet. Java was not designed to be source-code compatible with any other language. One outcome of this was a clean usable, pragmatic approach to objects. The object model in Java is simple and easy to extend, while simple types, such as integers, are kept as high-performance non-objects. Java architecture provides a portable, robust, high performing environment for development. Java provides portability by compiling the byte codes for the Java Virtual Machine, which is then interpreted on each platform by the run-time environment.

PYTHON

Python is a widely used high-level programming language for general-purpose programming, created by Guido van Rossum and first released in 1991. An interpreted language, Python has a design philosophy which emphasizes code readability and a syntax which allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale. Python features a dynamic type system and automatic memory management and supports multiple paradigms, including object-oriented, imperative, functional programming, and procedural styles. Python interpreters are available for many operating systems, allowing Python code to run on a wide variety of systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations.

BIG DATA:

Big data means really a big data; it is a collection of large datasets that cannot be processed using traditional computing techniques. Big data is not merely a data; rather it has become a complete subject, which involves various tools, techniques and frameworks. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business. To harness the power of big data, you would require an infrastructure that can manage and process huge volumes of structured and unstructured data in realtime and can protect data privacy and security.

HADOOP (TOOL USED):

Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation. Components include Hadoop Distributed File System (HDFS), which is capable of storing data across thousands of commodity servers to achieve high bandwidth between nodes; Hadoop Yet Another Resource Negotiator (YARN), which provides resource management and scheduling for user applications; and Hadoop MapReduce, which provides the programming model used to tackle large distributed data processing -- mapping data and reducing it to a result. MapReduce provides a new method of analysing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

APACHE PIG

Pig is a high-level platform for creating programs that run on Apache Hadoop. The language for this platform is called Pig Latin. Pig can execute its Hadoop jobs in Map Reduce, Apache Tez, or Apache Spark. Pig Latin abstracts the programming from the Java Map Reduce idiom into a notation which makes MapReduce programming high level, similar to that of SQL for RDBMSs. Pig Latin can be extended using User Defined Functions (UDFs) which the user can write in Java, Python, JavaScript, Ruby or Groovy, and then call directly from the language.

ECLIPSE:

Eclipse is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications, but it may also be used to develop applications in other Languages Development environments include the Eclipse Java development tools (JDT) for Java and Scala, Eclipse CDT for C/C++, and Eclipse PDT for PHP, among others.

6. MODULE DESCRIPTION:

1. Get symptoms Details.
   Collect the complete symptoms details of the patient and make it as a dataset. Here the dataset contains not only the symptoms but also the patient name, gender, age and their symptoms.
   HDFS is built to support applications with large data sets, including individual files that reach into the terabytes. So, run hadoop environment and then we can give the input file (log file) to HDFS i.e., the dataset are uploaded into HDFS.

2. K-mean clustering.
   K-means clustering is a method of vector quantization. It aims to partition n observations into k clusters in which each
observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Here K-mean clustering is used to cluster the symptoms of the patients.

3. Grouping.

Apply the grouping concept inorder to group the patient based on disease. Grouping can be done by using the clustered symptoms details of the patients which includes the parameters like patient name, age, gender and symptoms.

4. Predict disease.

Based on the result of grouping the Decision Support System (DSS) can be done. By examine a huge data and present the result of that examination and analysis to help users in making more effective and reliable business decisions.

7. EXPECTED OUTCOME:
Symptoms details in hdfs:

![File: /usr/Hadoop/dataset.csv](HDFS/usr/Hadoop/dataset.csv)
k-means clustering:

Grouping:
8. CONCLUSION:

In this paper we are proposing an idea to create such a CDSS clinical decision support system which will be useful for doctors to predict the diseases and user friendly and also to manage huge sets of data

9. REFERENCE


[4]. https://en.wikipedia.org/wiki/Clinical_decision_support_system clinical decision support system (CDSS) searchhealthit.techtarget.com › Decision support › Healthcare IT.