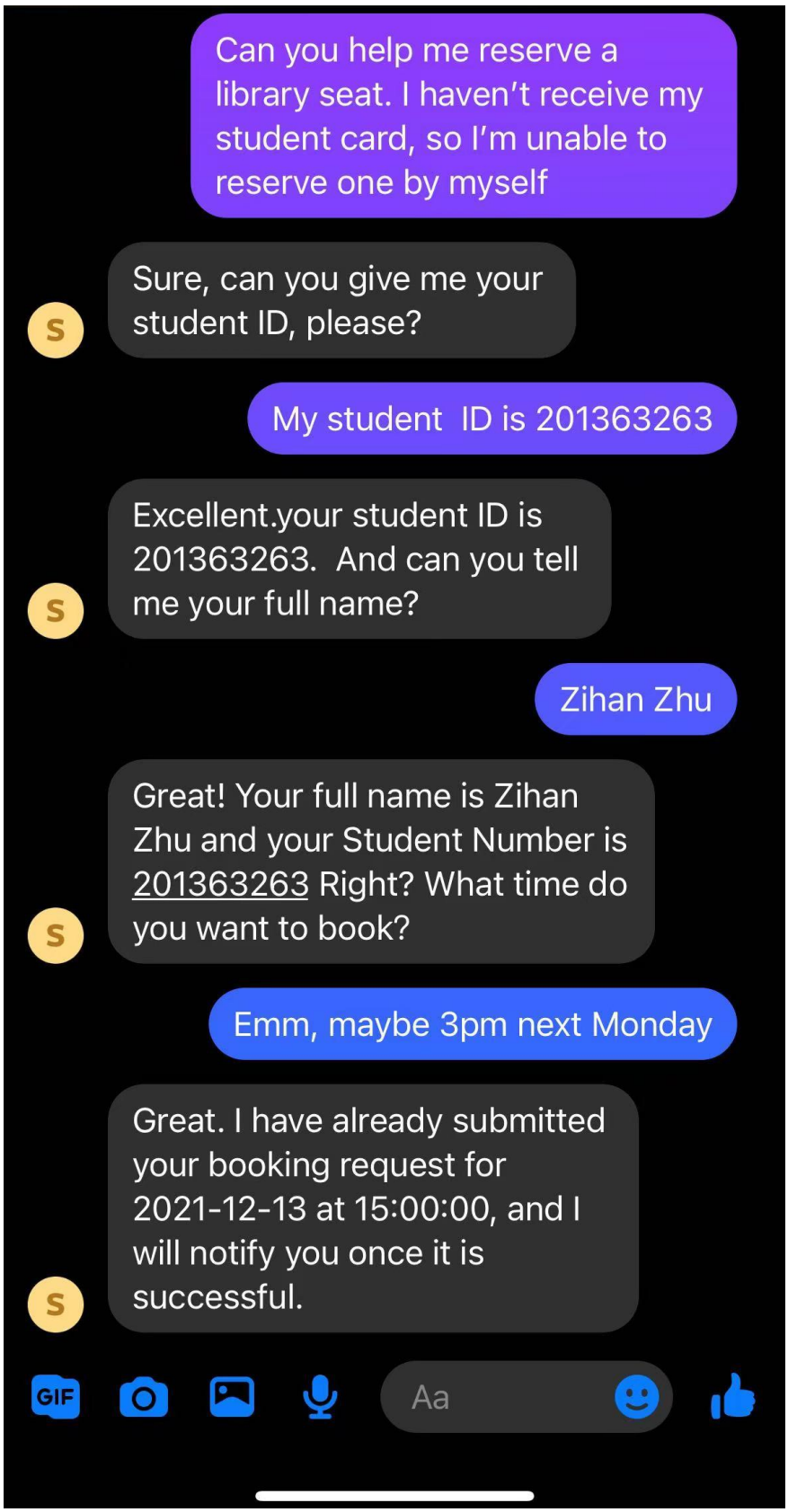
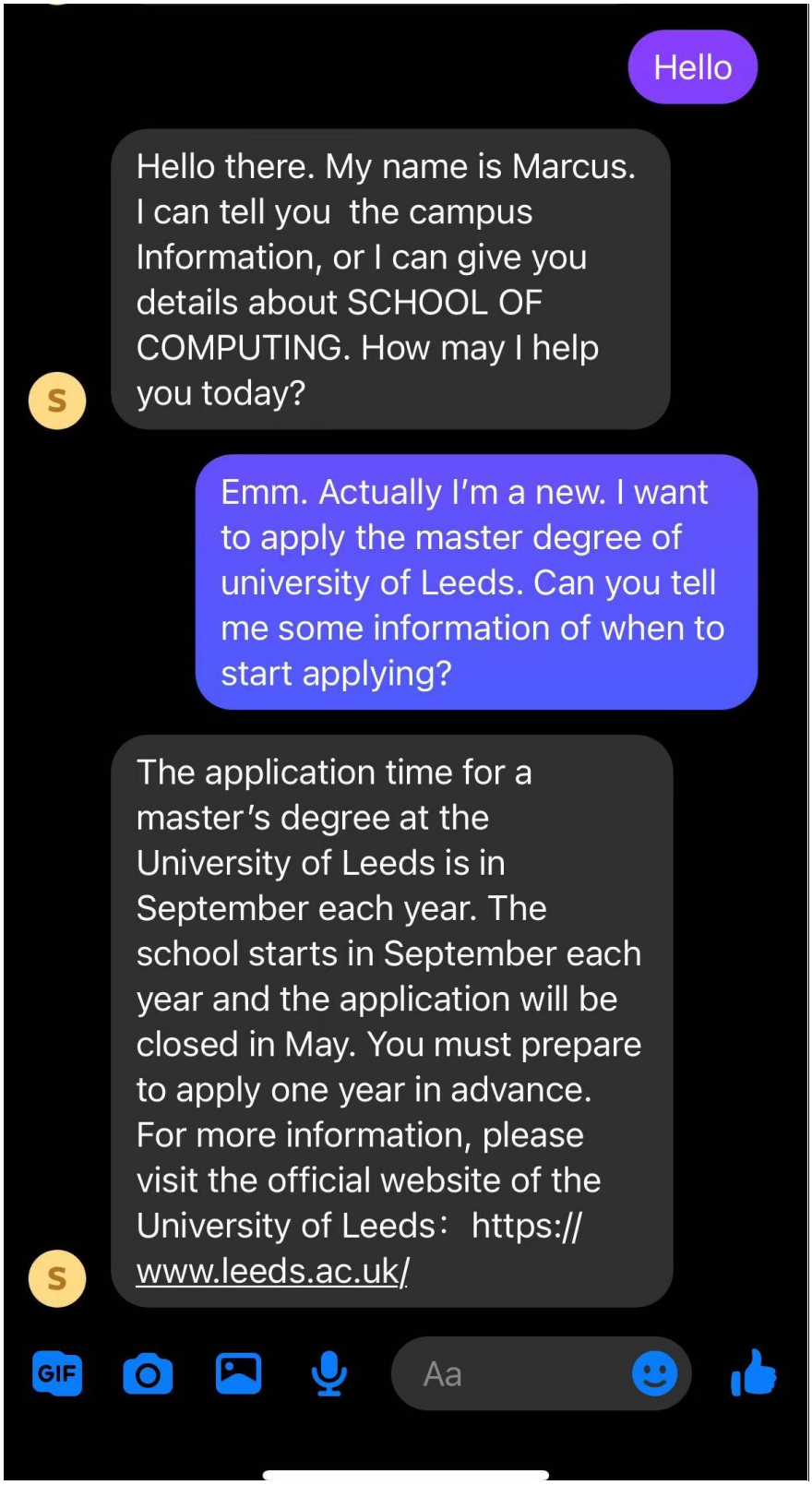
Library Reservation:

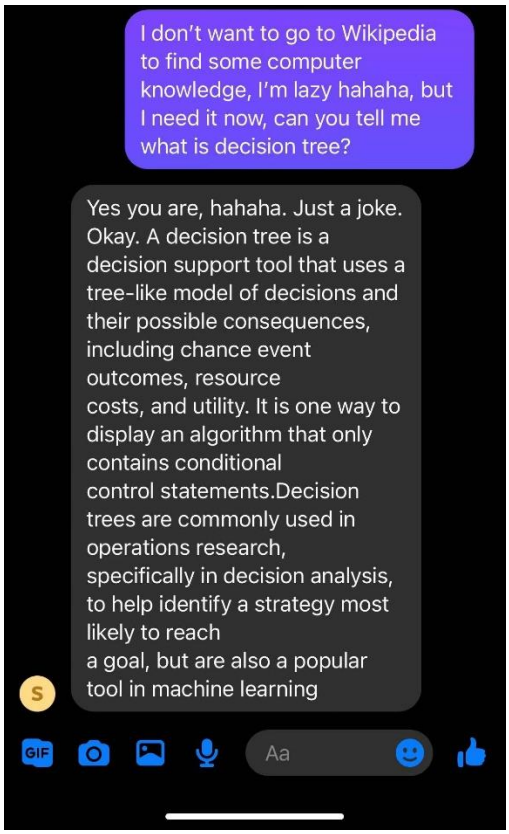


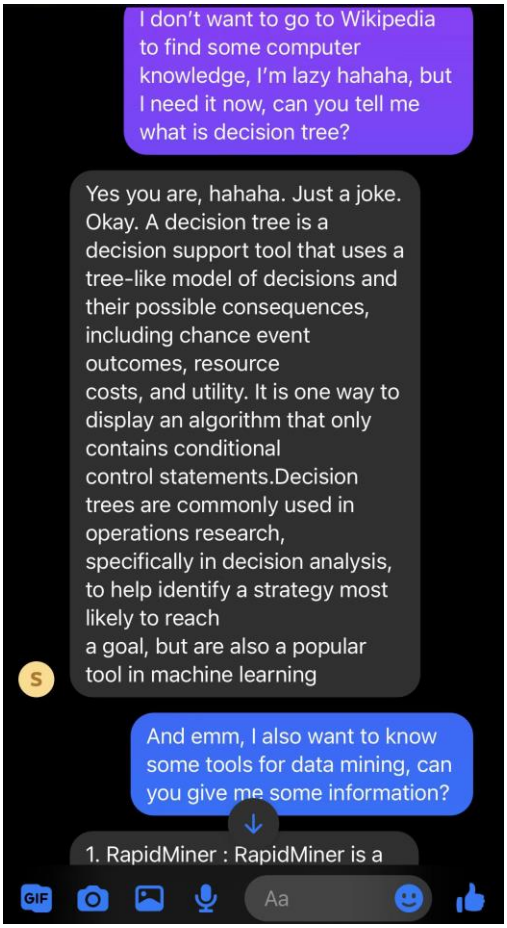
University Information Consultation:



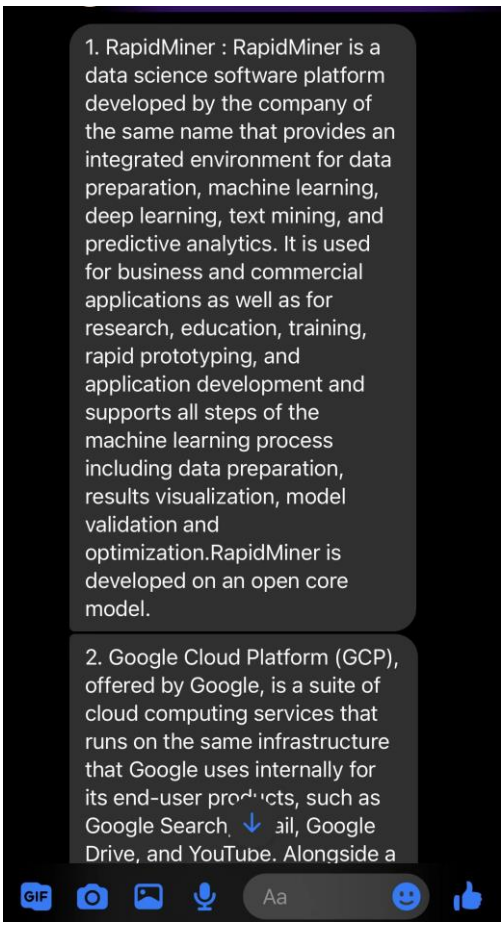
Question And Answer module:

Q1

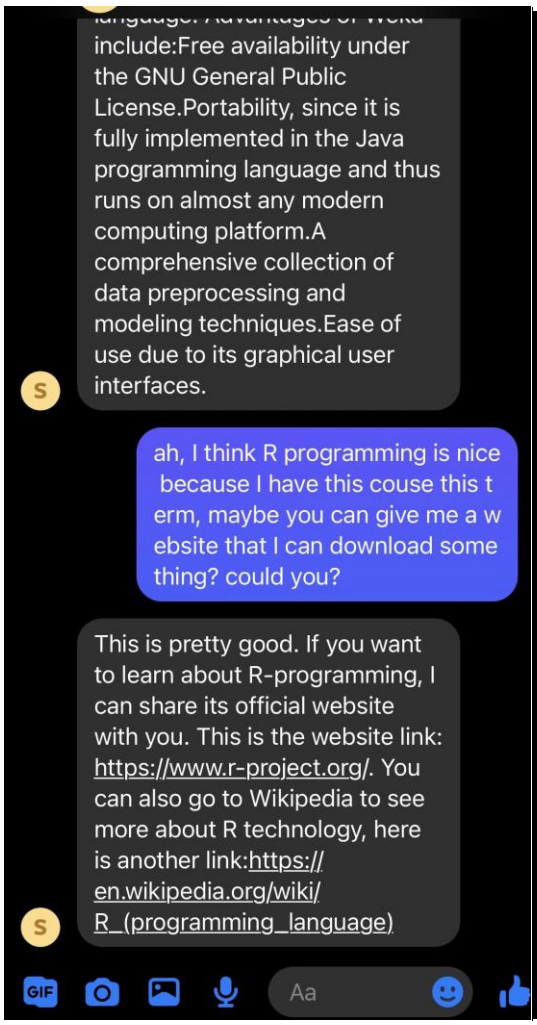



I don't want to go to Wikipedia to find some computer knowledge, I'm lazy hahaha, but I need it now, can you tell me what is decision tree?
Yes you are, hahaha. Just a joke. Okay. A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning
S
And emm, I also want to know some tools for data mining, can you give me some information?
1. RapidMiner : RapidMiner is a
GIF
Aa

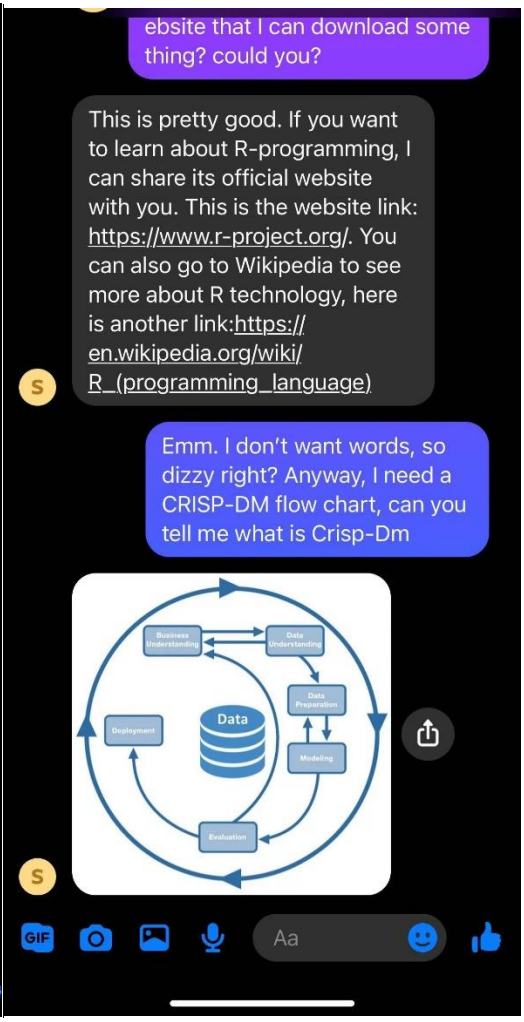
Q2


1. RapidMiner : RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.RapidMiner is developed on an open core model.
2. Google Cloud Platform (GCP), offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, Google Drive, and YouTube. Alongside a
GIF
Aa

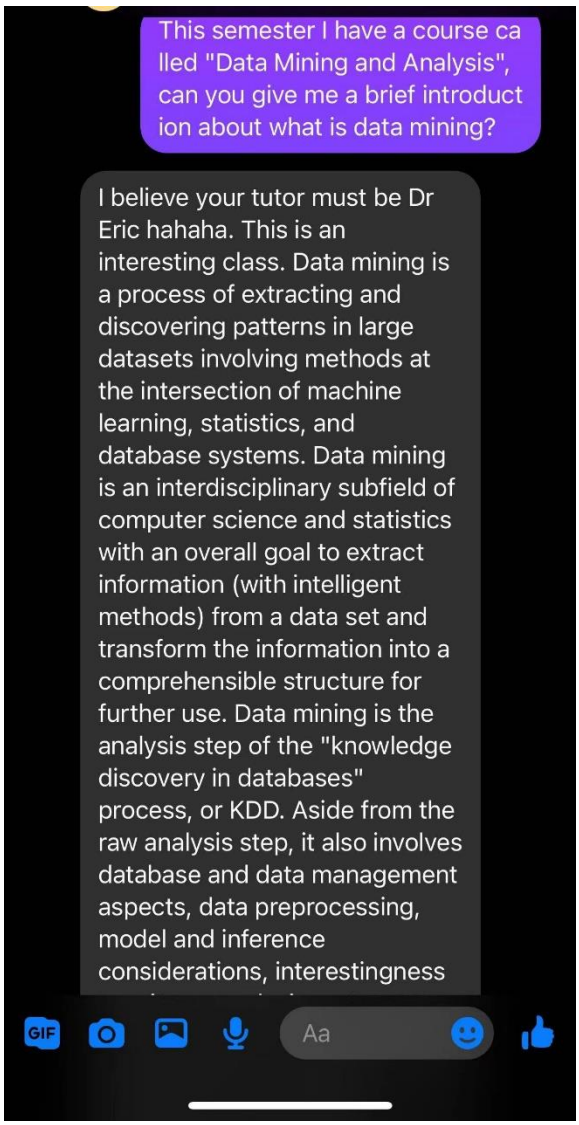
Q2


include:Free availability under the GNU General Public License.Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.A comprehensive collection of data preprocessing and modeling techniques.Ease of use due to its graphical user interfaces.
ah, I think R programming is nice because I have this couse this t erm, maybe you can give me a w ebsite that I can download some thing? could you?
This is pretty good. If you want to learn about R-programming, I can share its official website with you. This is the website link: https://www.r-project.org/. You can also go to Wikipedia to see more about R technology, here is another link:https://en.wikipedia.org/wiki/R_(programming_language)
GIF
Aa

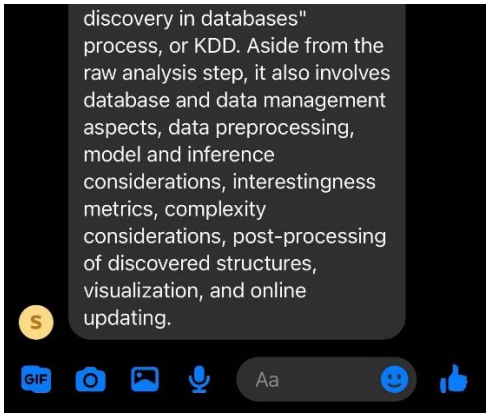
Q2


ebsite that I can download some thing? could you?
This is pretty good. If you want to learn about R-programming, I can share its official website with you. This is the website link: https://www.r-project.org/. You can also go to Wikipedia to see more about R technology, here is another link:https://en.wikipedia.org/wiki/R_(programming_language)
Emm. I don't want words, so dizzy right? Anyway, I need a CRISP-DM flow chart, can you tell me what is Crisp-Dm
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data
GIF
Aa

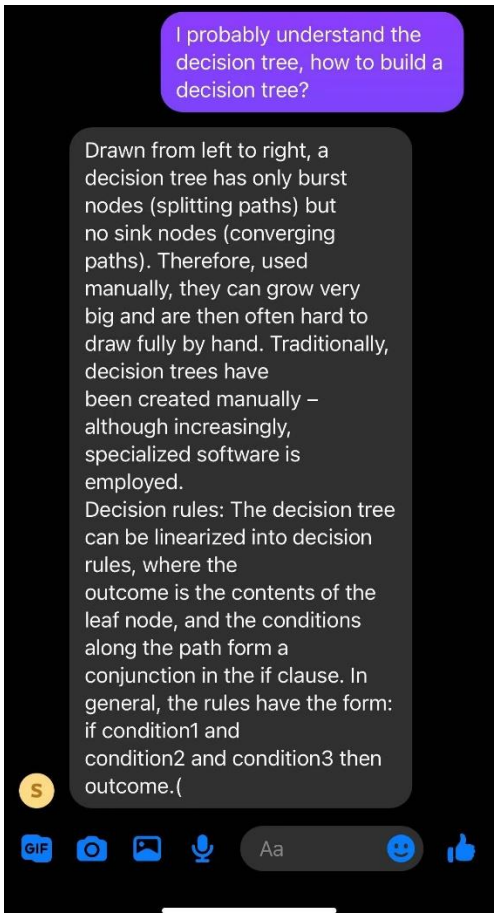
Q3


This semester I have a course ca lled "Data Mining and Analysis", can you give me a brief introduct ion about what is data mining?
I believe your tutor must be Dr Eric hahaha. This is an interesting class. Data mining is a process of extracting and discovering patterns in large datasets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data preprocessing, model and inference considerations, interestingness
GIF
Aa

Q4


discovery in databases" process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.
S
GIF
Aa

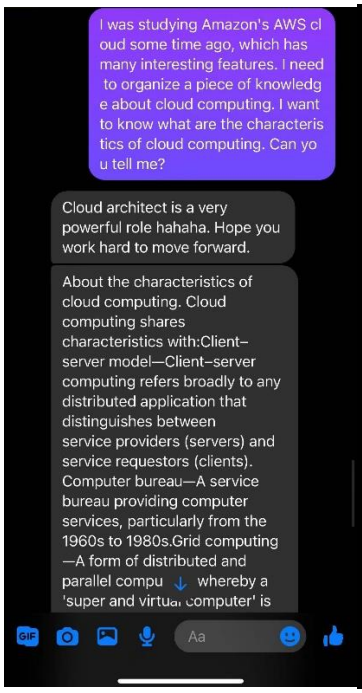
Q4

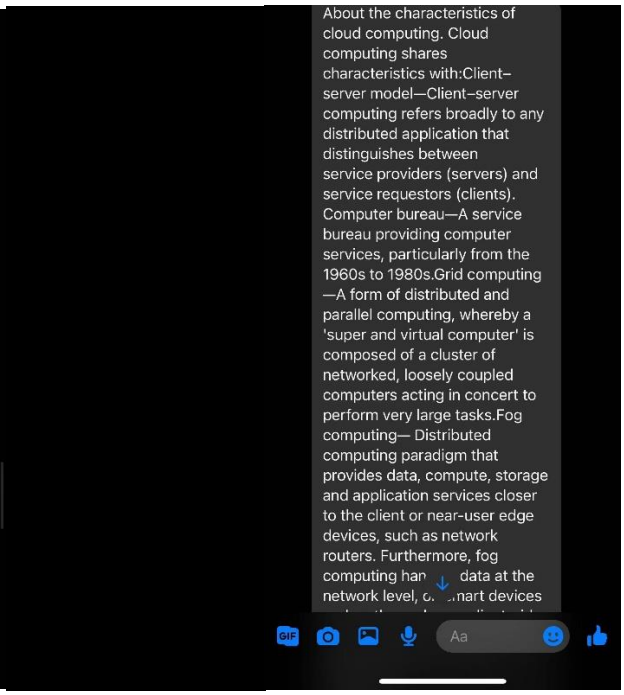

I probably understand the decision tree, how to build a decision tree?
Drawn from left to right, a decision tree has only burst nodes (splitting paths) but no sink nodes (converging paths). Therefore, used manually, they can grow very big and are then often hard to draw fully by hand. Traditionally, decision trees have been created manually – although increasingly, specialized software is employed.
Decision rules: The decision tree can be linearized into decision rules, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form: if condition1 and condition2 and condition3 then outcome.(
S
GIF
Aa

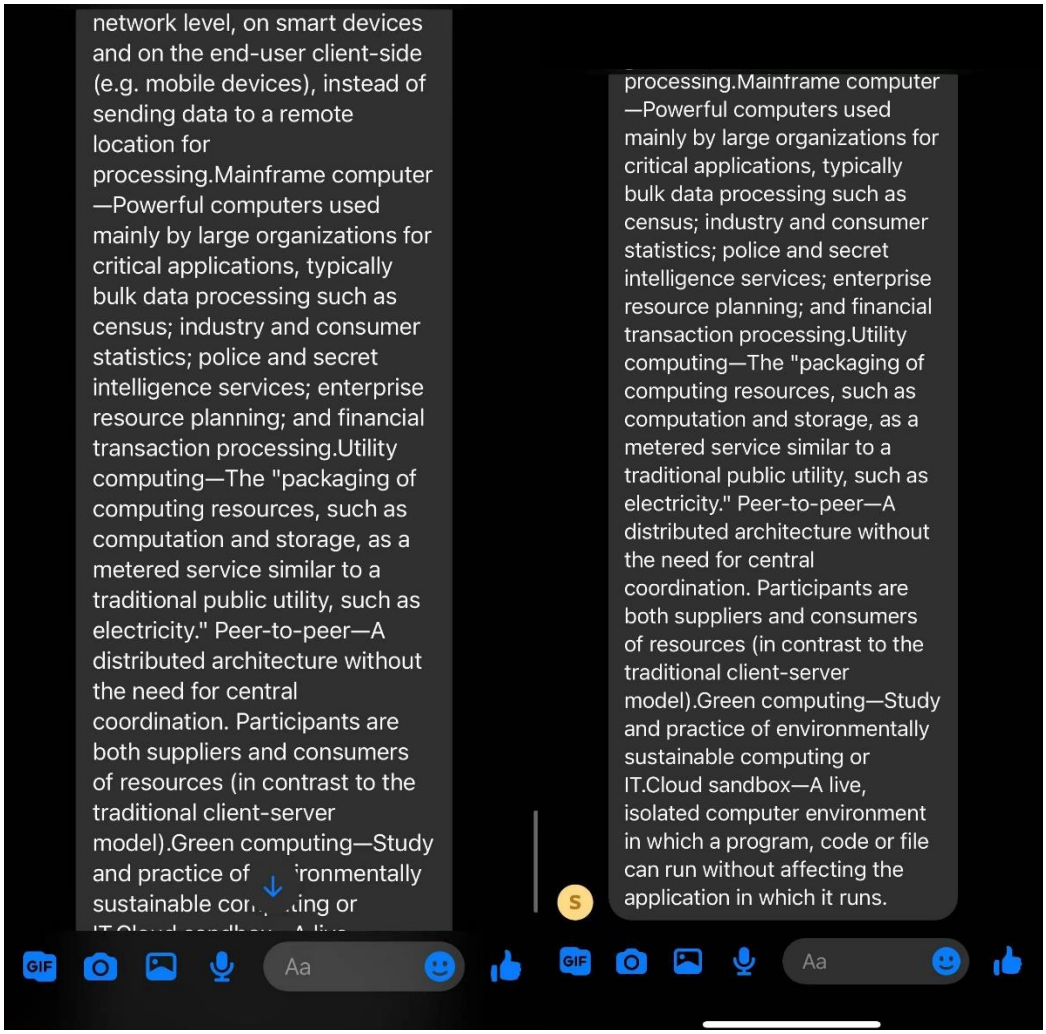
Q5

Q6

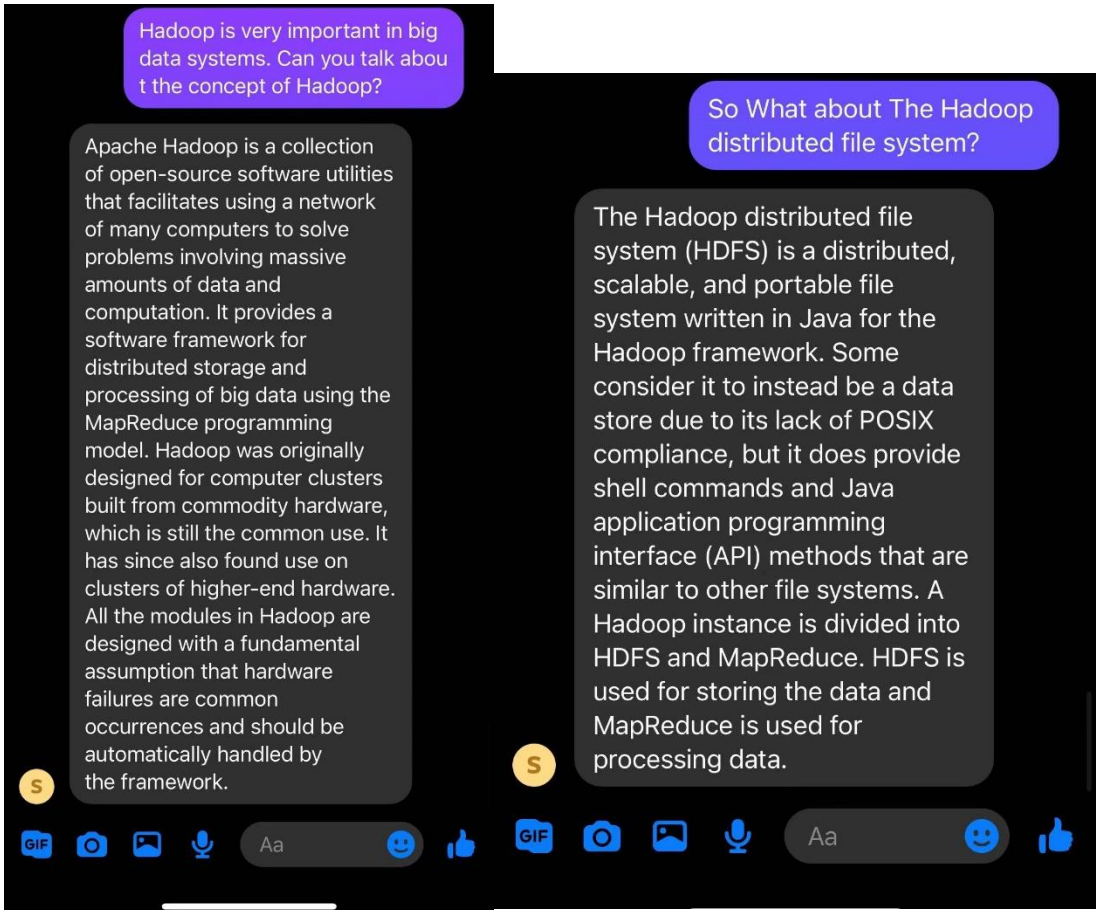
Q6


I was studying Amazon's AWS cl oud some time ago, which has many interesting features. I need to organize a piece of knowledg e about cloud computing. I want to know what are the characteris tics of cloud computing. Can yo u tell me?
Cloud architect is a very powerful role hahaha. Hope you work hard to move forward.
About the characteristics of cloud computing. Cloud computing shares characteristics with:Client–server model—Client–server computing refers broadly to any distributed application that distinguishes between service providers (servers) and service requestors (clients). Computer bureau—A service bureau providing computer services, particularly from the 1960s to 1980s.Grid computing —A form of distributed and parallel compu whereby a 'super and virtua computer' is
GIF
Aa

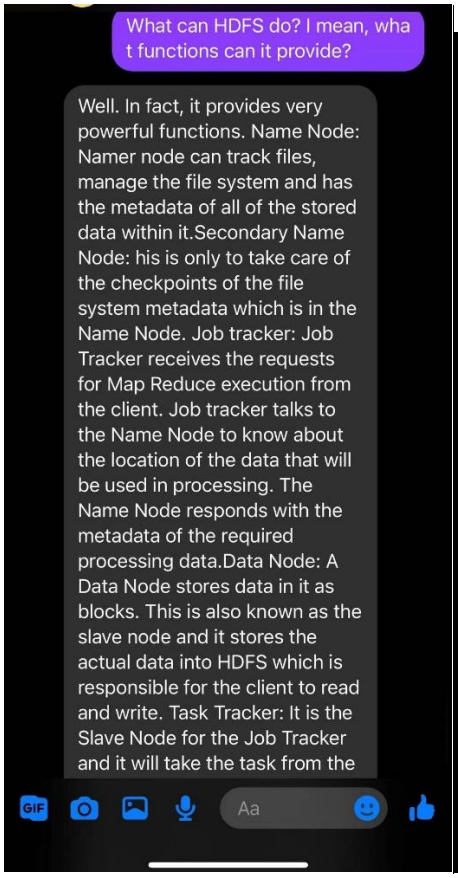

About the characteristics of cloud computing. Cloud computing shares characteristics with:Client–server model—Client–server computing refers broadly to any distributed application that distinguishes between service providers (servers) and service requestors (clients). Computer bureau—A service bureau providing computer services, particularly from the 1960s to 1980s.Grid computing —A form of distributed and parallel computing, whereby a 'super and virtual computer' is composed of a cluster of networked, loosely coupled computers acting in concert to perform very large tasks.Fog computing— Distributed computing paradigm that provides data, compute, storage and application services closer to the client or near-user edge devices, such as network routers. Furthermore, fog computing han data at the network level,
GIF
Aa


network level, on smart devices and on the end-user client-side (e.g. mobile devices), instead of sending data to a remote location for processing.Mainframe computer—Powerful computers used mainly by large organizations for critical applications, typically bulk data processing such as census; industry and consumer statistics; police and secret intelligence services; enterprise resource planning; and financial transaction processing.Utility computing—The "packaging of computing resources, such as computation and storage, as a metered service similar to a traditional public utility, such as electricity." Peer-to-peer—A distributed architecture without the need for central coordination. Participants are both suppliers and consumers of resources (in contrast to the traditional client-server model).Green computing—Study and practice of
ironmentally sustainable cor
ing or
GIF
Aa
processing.Mainframe computer—Powerful computers used mainly by large organizations for critical applications, typically bulk data processing such as census; industry and consumer statistics; police and secret intelligence services; enterprise resource planning; and financial transaction processing.Utility computing—The "packaging of computing resources, such as computation and storage, as a metered service similar to a traditional public utility, such as electricity." Peer-to-peer—A distributed architecture without the need for central coordination. Participants are both suppliers and consumers of resources (in contrast to the traditional client-server model).Green computing—Study and practice of environmentally sustainable computing or IT.Cloud sandbox—A live, isolated computer environment in which a program, code or file can run without affecting the application in which it runs.
S
GIF
Aa

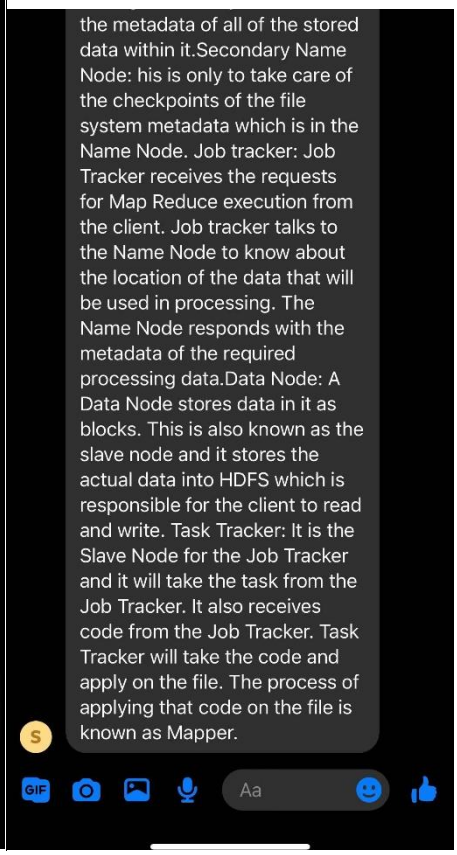
Q6 **Q6**


Hadoop is very important in big data systems. Can you talk abou t the concept of Hadoop?
Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. Hadoop was originally designed for computer clusters built from commodity hardware, which is still the common use. It has since also found use on clusters of higher-end hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework.
S
GIF
Aa
So What about The Hadoop distributed file system?
The Hadoop distributed file system (HDFS) is a distributed, scalable, and portable file system written in Java for the Hadoop framework. Some consider it to instead be a data store due to its lack of POSIX compliance, but it does provide shell commands and Java application programming interface (API) methods that are similar to other file systems. A Hadoop instance is divided into HDFS and MapReduce. HDFS is used for storing the data and MapReduce is used for processing data.
S
GIF
Aa

Q7 **Q8**

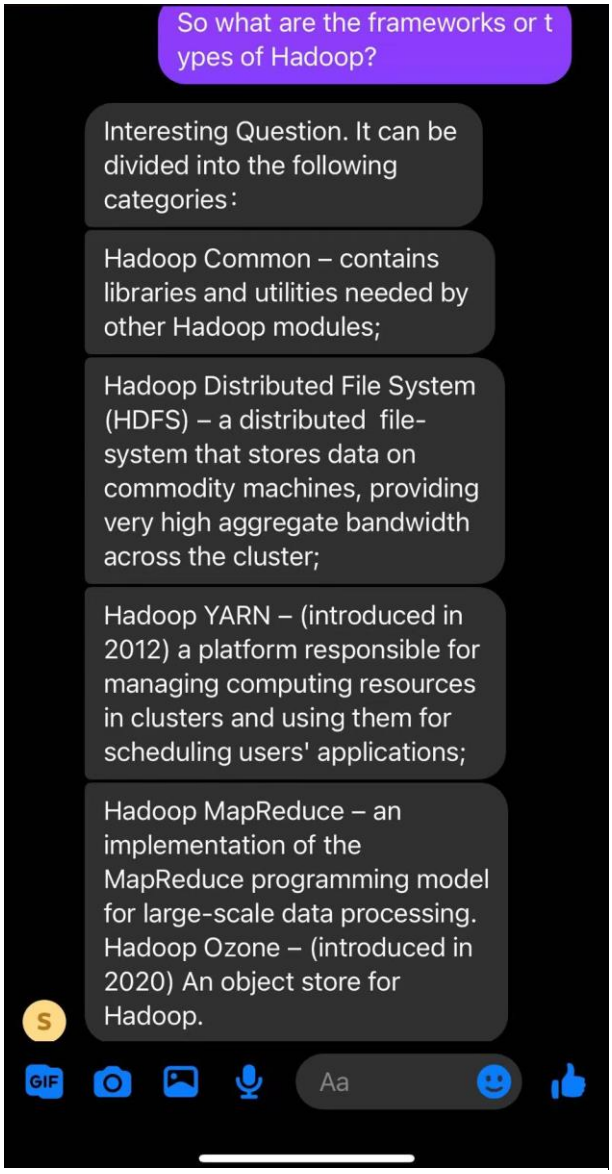

What can HDFS do? I mean, wha t functions can it provide?
Well. In fact, it provides very powerful functions. Name Node: Namer node can track files, manage the file system and has the metadata of all of the stored data within it.Secondary Name Node: his is only to take care of the checkpoints of the file system metadata which is in the Name Node. Job tracker: Job Tracker receives the requests for Map Reduce execution from the client. Job tracker talks to the Name Node to know about the location of the data that will be used in processing. The Name Node responds with the metadata of the required processing data.Data Node: A Data Node stores data in it as blocks. This is also known as the slave node and it stores the actual data into HDFS which is responsible for the client to read and write. Task Tracker: It is the Slave Node for the Job Tracker and it will take the task from the
GIF
Aa

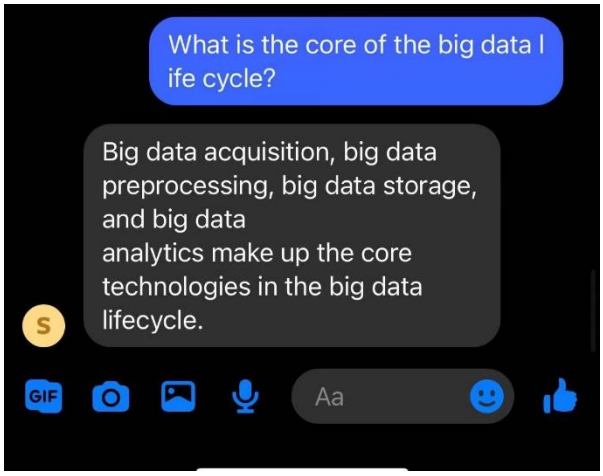
Q9


the metadata of all of the stored data within it.Secondary Name Node: his is only to take care of the checkpoints of the file system metadata which is in the Name Node. Job tracker: Job Tracker receives the requests for Map Reduce execution from the client. Job tracker talks to the Name Node to know about the location of the data that will be used in processing. The Name Node responds with the metadata of the required processing data.Data Node: A Data Node stores data in it as blocks. This is also known as the slave node and it stores the actual data into HDFS which is responsible for the client to read and write. Task Tracker: It is the Slave Node for the Job Tracker and it will take the task from the Job Tracker. It also receives code from the Job Tracker. Task Tracker will take the code and apply on the file. The process of applying that code on the file is known as Mapper.
S
GIF
Aa

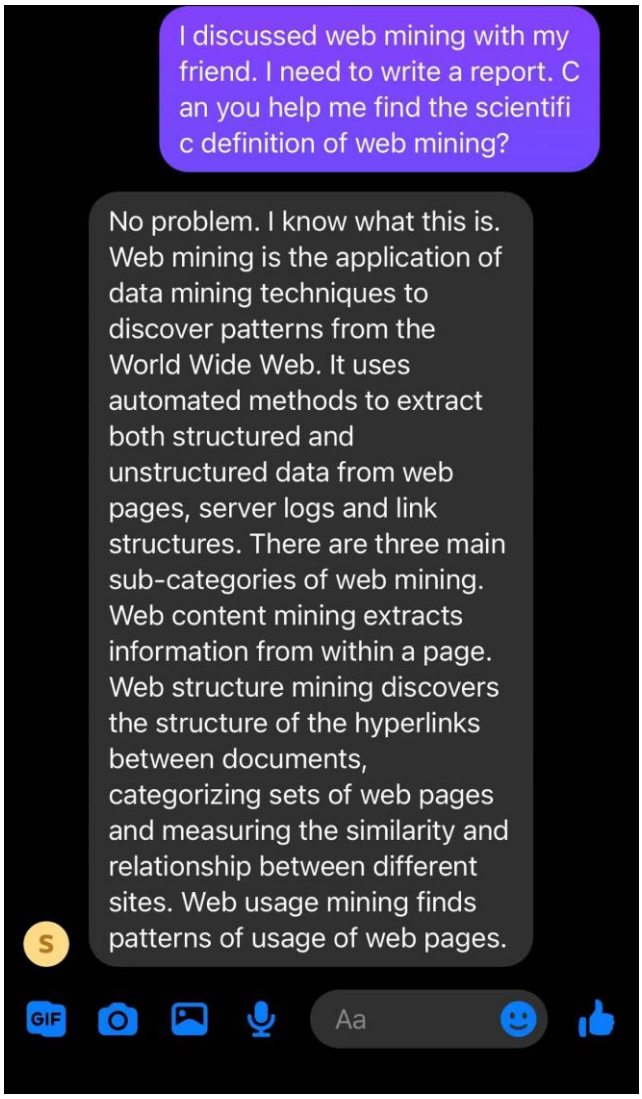
Q9

Q10


So what are the frameworks or t ypes of Hadoop?
Interesting Question. It can be divided into the following categories:
Hadoop Common – contains libraries and utilities needed by other Hadoop modules;
Hadoop Distributed File System (HDFS) – a distributed file-system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster;
Hadoop YARN – (introduced in 2012) a platform responsible for managing computing resources in clusters and using them for scheduling users' applications;
Hadoop MapReduce – an implementation of the MapReduce programming model for large-scale data processing.
Hadoop Ozone – (introduced in 2020) An object store for Hadoop.
S
GIF
Aa


What is the core of the big data l ife cycle?
Big data acquisition, big data preprocessing, big data storage, and big data
analytics make up the core technologies in the big data lifecycle.
S
GIF
Aa

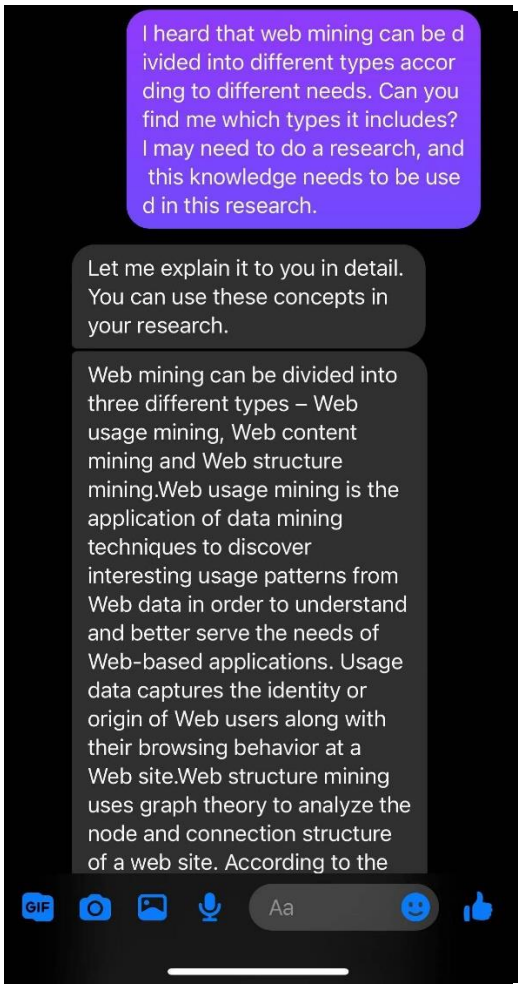
Q11


I discussed web mining with my friend. I need to write a report. C an you help me find the scientifi c definition of web mining?
No problem. I know what this is. Web mining is the application of data mining techniques to discover patterns from the World Wide Web. It uses automated methods to extract both structured and unstructured data from web pages, server logs and link structures. There are three main sub-categories of web mining. Web content mining extracts information from within a page. Web structure mining discovers the structure of the hyperlinks between documents, categorizing sets of web pages and measuring the similarity and relationship between different sites. Web usage mining finds patterns of usage of web pages.
S
GIF
Aa

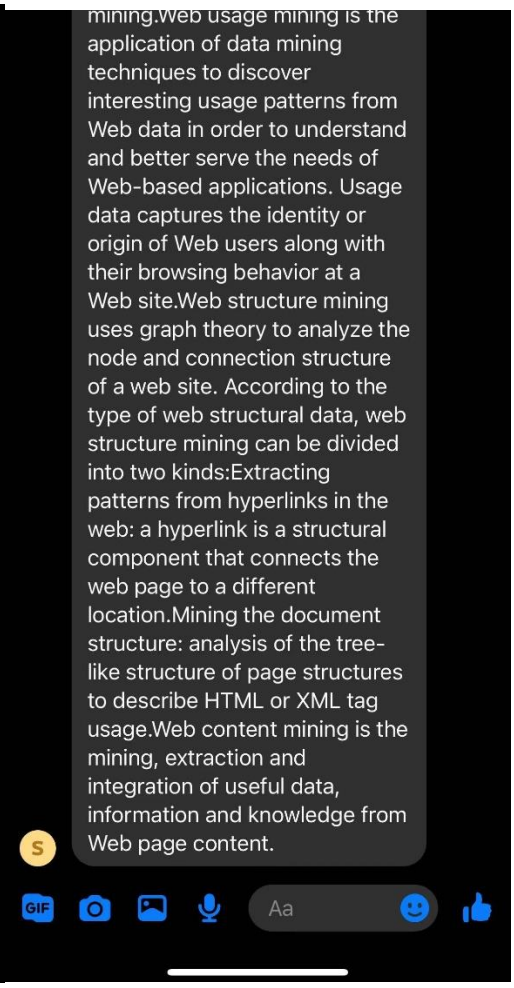
Q12


What can be gained by using we b mining technology?
Simple question. Get competitor and customer information. Discover user access patterns. Anticompetitive intelligence activities
S
GIF
Aa

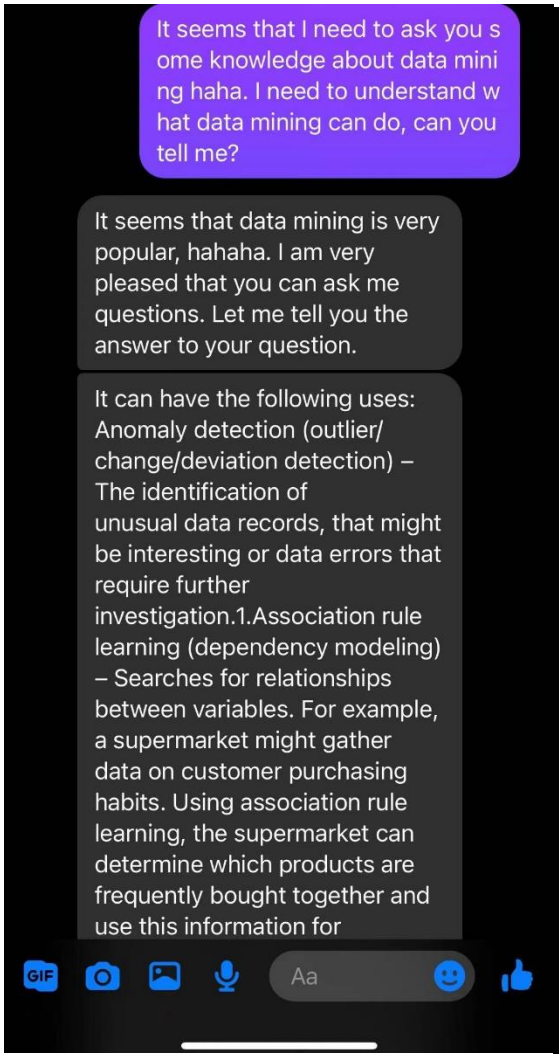
Q13


I heard that web mining can be d ivided into different types accor ding to different needs. Can you find me which types it includes? I may need to do a research, and this knowledge needs to be use d in this research.
Let me explain it to you in detail. You can use these concepts in your research.
Web mining can be divided into three different types – Web usage mining, Web content mining and Web structure mining.Web usage mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.Web structure mining uses graph theory to analyze the node and connection structure of a web site. According to the
GIF
Aa

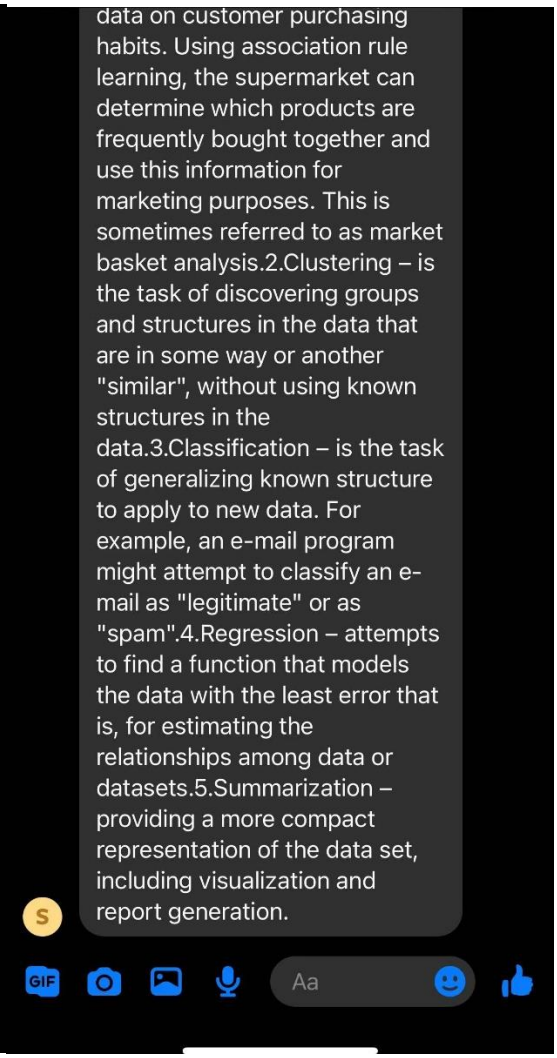
Q14


mining.Web usage mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.Web structure mining uses graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page content.
S
GIF
Aa

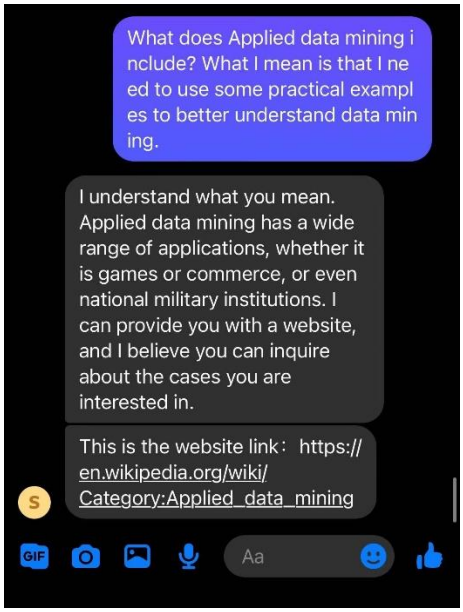
Q14


It seems that I need to ask you s ome knowledge about data mini ng haha. I need to understand w hat data mining can do, can you tell me?
It seems that data mining is very popular, hahaha. I am very pleased that you can ask me questions. Let me tell you the answer to your question.
It can have the following uses: Anomaly detection (outlier/ change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation.1.Association rule learning (dependency modeling) – Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for
GIF
Aa

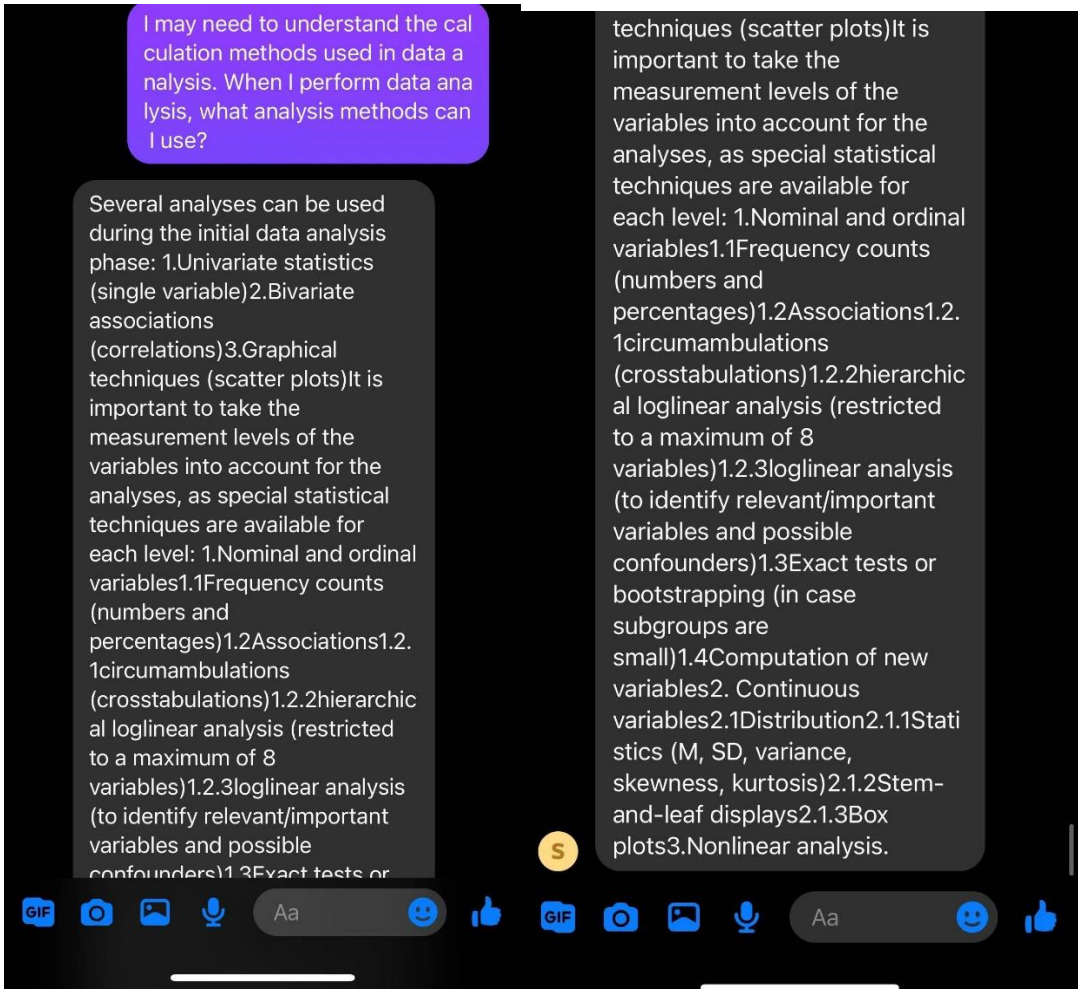
Q15


data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.2.Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.3.Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".4.Regression – attempts to find a function that models the data with the least error that is, for estimating the relationships among data or datasets.5.Summarization – providing a more compact representation of the data set, including visualization and report generation.
S
GIF
Aa

Q15

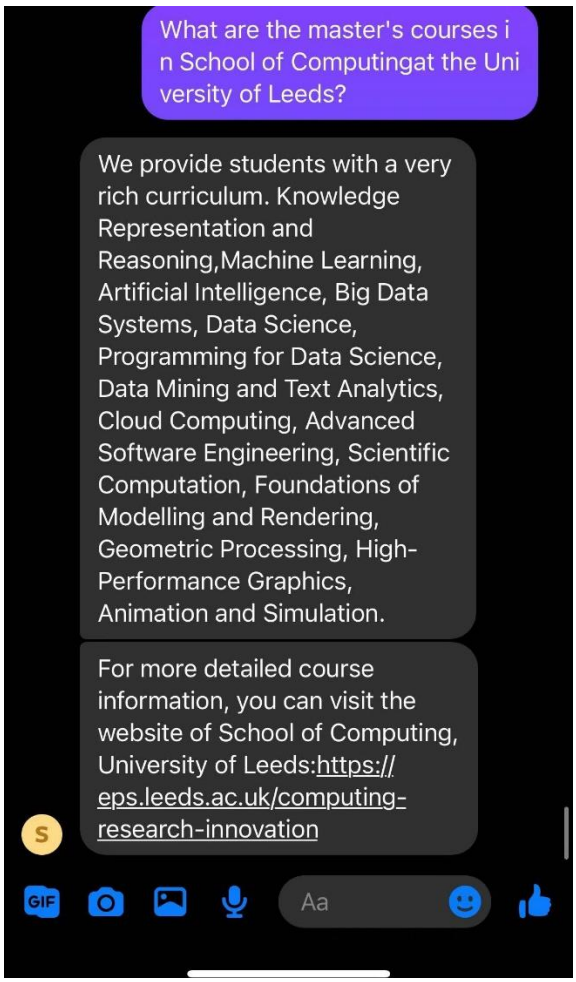

What does Applied data mining i nclude? What I mean is that I ne ed to use some practical exampl es to better understand data min ing.
I understand what you mean. Applied data mining has a wide range of applications, whether it is games or commerce, or even national military institutions. I can provide you with a website, and I believe you can inquire about the cases you are interested in.
This is the website link: https:// en.wikipedia.org/wiki/ Category:Applied_data_mining
S
GIF
Aa

Q16

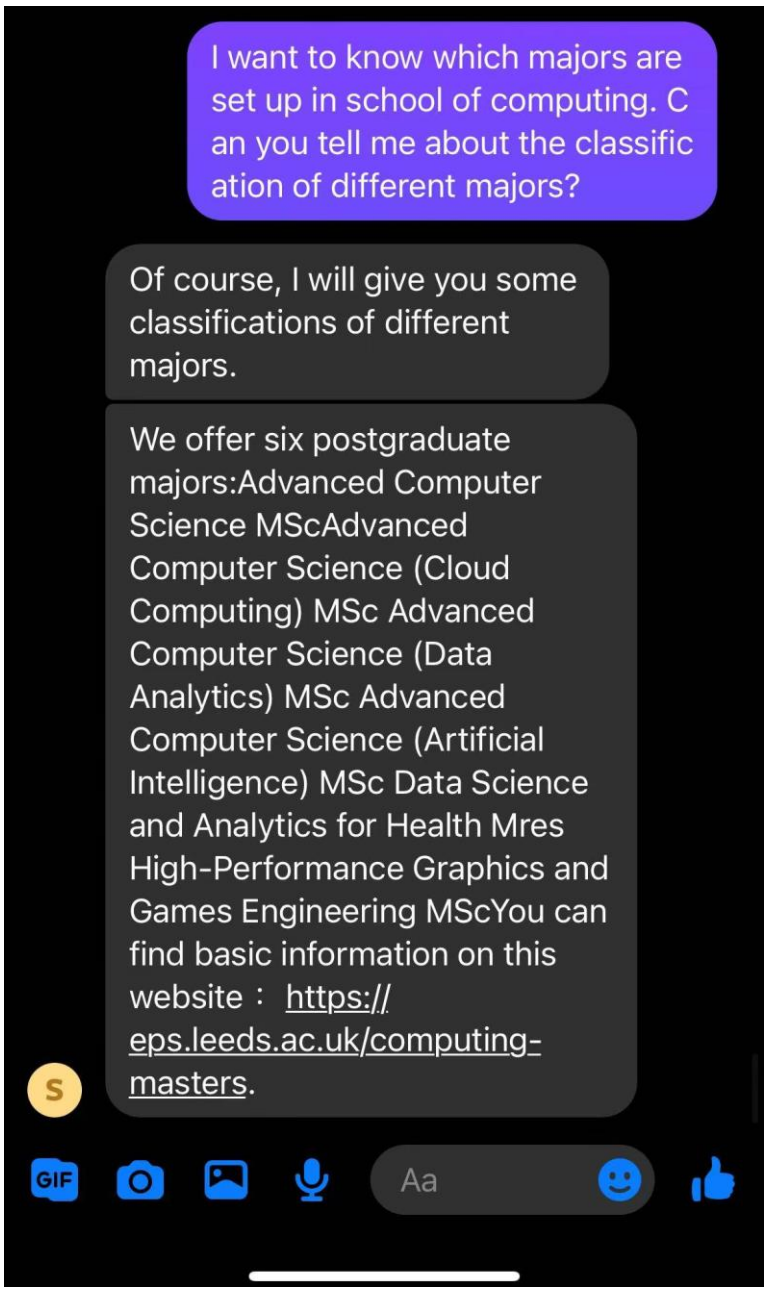

I may need to understand the cal culation methods used in data a nalysis. When I perform data ana lysis, what analysis methods can I use?
Several analyses can be used during the initial data analysis phase: 1.Univariate statistics (single variable)2.Bivariate associations (correlations)3.Graphical techniques (scatter plots)It is important to take the measurement levels of the variables into account for the analyses, as special statistical techniques are available for each level: 1.Nominal and ordinal variables1.1Frequency counts (numbers and percentages)1.2Associations1.2.1circumambulations (crosstabulations)1.2.2hierarchical loglinear analysis (restricted to a maximum of 8 variables)1.2.3loglinear analysis (to identify relevant/important variables and possible confounders)1.3Exact tests or
techniques (scatter plots)It is important to take the measurement levels of the variables into account for the analyses, as special statistical techniques are available for each level: 1.Nominal and ordinal variables1.1Frequency counts (numbers and percentages)1.2Associations1.2.1circumambulations (crosstabulations)1.2.2hierarchical loglinear analysis (restricted to a maximum of 8 variables)1.2.3loglinear analysis (to identify relevant/important variables and possible confounders)1.3Exact tests or bootstrapping (in case subgroups are small)1.4Computation of new variables2. Continuous variables2.1Distribution2.1.1Statistics (M, SD, variance, skewness, kurtosis)2.1.2Stem-and-leaf displays2.1.3Box plots3.Nonlinear analysis.
S
GIF
Aa
GIF
Aa

Q17

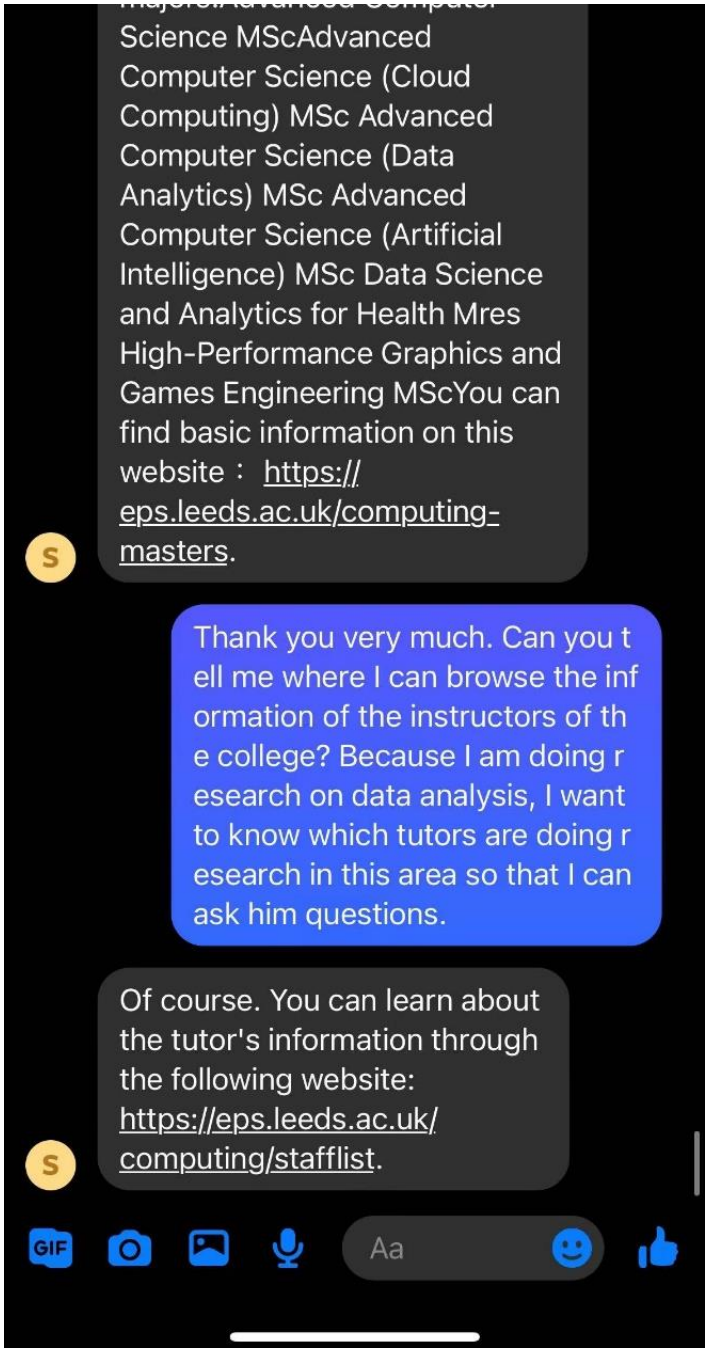
Q17


What are the master's courses i n School of Computingat the Uni versity of Leeds?
We provide students with a very rich curriculum. Knowledge Representation and Reasoning,Machine Learning, Artificial Intelligence, Big Data Systems, Data Science, Programming for Data Science, Data Mining and Text Analytics, Cloud Computing, Advanced Software Engineering, Scientific Computation, Foundations of Modelling and Rendering, Geometric Processing, High-Performance Graphics, Animation and Simulation.
For more detailed course information, you can visit the website of School of Computing, University of Leeds:https://eps.leeds.ac.uk/computing-research-innovation
S
GIF
Aa

Q18


I want to know which majors are set up in school of computing. C an you tell me about the classific ation of different majors?
Of course, I will give you some classifications of different majors.
We offer six postgraduate majors:Advanced Computer Science MScAdvanced Computer Science (Cloud Computing) MSc Advanced Computer Science (Data Analytics) MSc Advanced Computer Science (Artificial Intelligence) MSc Data Science and Analytics for Health Mres High-Performance Graphics and Games Engineering MScYou can find basic information on this website： https:// eps.leeds.ac.uk/computing-masters.
S
GIF
Aa

Q19


Science MScAdvanced Computer Science (Cloud Computing) MSc Advanced Computer Science (Data Analytics) MSc Advanced Computer Science (Artificial Intelligence) MSc Data Science and Analytics for Health Mres High-Performance Graphics and Games Engineering MScYou can find basic information on this website： https://eps.leeds.ac.uk/computing-masters.
S
Thank you very much. Can you t ell me where I can browse the inf ormation of the instructors of th e college? Because I am doing r esearch on data analysis, I want to know which tutors are doing r esearch in this area so that I can ask him questions.
Of course. You can learn about the tutor's information through the following website: https://eps.leeds.ac.uk/computing/stafflist.
S
GIF
Aa

Q20