Experiment No.10

A.1 Aim: Study of complex Data Types used in Web mining and Spatial Mining Applications.

PART B

(PART B: TO BE COMPLETED BY STUDENTS)

(Students must submit the soft copy as per the following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Blackboard access available)

Roll No. 50	**Name:** AMEY THAKUR
Class: Comps TE B	**Batch:** B3
Date of Experiment: 28/04/2021	**Date of Submission:** 28/04/2021
Grade:	

B.1 Software Code written by a student:

(Paste your problem statement related to your case study completed during the 2 hours of practice in the lab here)

Web Mining:

Web mining is an application of data mining techniques to find information patterns from web data. Web mining helps to improve the power of web search engines by identifying web pages and classifying web documents. Web mining is very useful for e-commerce websites and e-services.

Spatial Data Mining:

Spatial data mining is the application of data mining to spatial models. In spatial data mining, analysts use geographical or spatial information to produce business intelligence or other results. This requires specific techniques and resources to get geographical data into relevant and useful formats.

B.2 Input and Output:

(Paste your program input and output in the following format, If there is an error then paste the specific error in the output part. In case of an error with the due permission of the faculty, an extension can be given to submit the error-free code with output in due course of time. Students will be graded accordingly.)

Web mining can be broadly divided into three categories:

1. Web Content Mining
2. Web Structure Mining
3. Web Usage Mining

1. Web Content Mining

➔ Web content mining targets knowledge discovery, in which the main objects are the traditional collections of multimedia documents such as images, video, and audio, which are embedded in or linked to the web pages.

➔ It is also quite different from Data mining because Web data are mainly semi-structured and/or unstructured, while Data mining deals primarily with structured data. Web content mining is different because of the semi-structured nature of the Web.

➔ Web content mining could be differentiated from two points of view: Agent-based approach or Database approach. The first approach aims at improving information finding and filtering. The second approach aims at modelling the data on the Web into a more structured form to apply a standard database querying mechanism and data mining applications to analyze it.

2. Web Structure Mining

➔ Web Structure Mining focuses on the analysis of the link structure of the web and one of its purposes is to identify more preferable documents. The different objects are linked in some way. The intuition is that a hyperlink from document A to document B implies that the author of the document. A thinks document B contains worthwhile information. Web structure mining helps in discovering similarities between web sites or discovering important sites for a particular topic or discipline or in discovering web communities.

➔ Simply applying the traditional processes and assuming that the events are independent can lead to wrong conclusions. However, the appropriate handling of the links could lead to potential correlations and then improve the predictive accuracy of the learned models.

➔ The goal of Web structure mining is to generate a structural summary of the Web site and Web page. Technically, Web content mining mainly focuses on the structure of inner-document, while Web structure mining tries to discover the link structure of the hyperlinks at the inter-document level. Based on the topology of the hyperlinks, Web structure mining will

categorize the Web pages and generate information, such as the similarity and relationship between different Web sites.

3. **Web Usage Mining**

➔ Web Usage Mining focuses on techniques that could predict the behaviour of users while they are interacting with the WWW. Web usage mining, discovering user navigation patterns from web data, try to discover the useful information from the secondary data derived from the interactions of the users while surfing on the Web. Web usage mining collects the data from Weblog records to discover user access patterns of web pages. There are several available research projects and commercial tools that analyze those patterns for different purposes. The insight knowledge could be utilized in personalization, system improvement, site modification, business intelligence and usage characterization.

➔ The only information left behind by many users visiting a Website is the path through the pages they have accessed. Most of the Web information retrieval tools only use textual information, while they ignore the link information that could be very valuable.

B.3 Observations and learning:

(Students are expected to comment on the output obtained with clear observations and learning for each task/ subpart assigned)

➔ Web mining helps to improve the power of web search engines by classifying web documents and identifying the web pages.

➔ It is used for Web Searching e.g., Google, Yahoo etc. and Vertical Searching e.g., FatLens, Become etc.

➔ Web mining is used to predict user behaviour.

➔ Web mining is beneficial for a particular Website and e-service, e.g., landing page optimization.

➔ Web mining can be broadly divided into three different types of techniques of mining: Web Content Mining, Web Structure Mining, and Web Usage Mining.

B.4 Conclusion:

(Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)

We understood the applications of web mining and spatial mining. We learnt the types of web mining and the concept of web mining.

B.5 Question of Curiosity

(To be answered by the student based on the practical performed and learning/observations)

1. Explain Mining of Spatial Association and colocation Pattern.

Ans:

(1) **Spatial collaboration:** Spatial collocation patterns associate the co-existence of non-spatial features in a spatial neighbourhood. An example of such a pattern can associate contaminated water reservoirs with certain diseases in their spatial neighbourhood. Previous work on discovering collocation patterns converts neighbourhoods of feature instances to itemsets and applies mining techniques for transactional data to discover the patterns. We propose a method that combines the discovery of spatial neighbourhoods with the mining process. Our technique is an extension of a spatial join algorithm that operates on multiple inputs and counts long pattern instances. As demonstrated by experimentation, it yields significant performance improvements compared to previous approaches.

(2) **Spatial association:** Spatial association is the degree to which things are similarly arranged in space. Analysis of the distribution patterns of two phenomena is done by map overlay. If the distributions are similar, then the spatial association is strong, and vice versa. In a Geographic Information System, the analysis can be done quantitatively. For example, a set of observations (as points or extracted from raster cells) at matching locations can be intersected and examined by regression analysis. Like spatial autocorrelation, this can be a useful tool for spatial prediction. In spatial modelling, the concept of spatial association allows the use of covariates in a regression equation to predict the geographic field and thus produce a map.

2. Write down Applications of web mining in detail?

Ans:

Financial Data Analysis:

The financial data in the banking and financial industry is generally reliable and of high quality which facilitates systematic data analysis and data mining. Some of the typical cases are as follows –

- Design and construction of data warehouses for multidimensional data analysis and data mining.
- Loan payment prediction and customer credit policy analysis.
- Classification and clustering of customers for targeted marketing.
- Detection of money laundering and other financial crimes.

Retail Industry:

Data Mining has its great application in the Retail Industry because it collects a large amount of data from sales, customer purchasing history, goods transportation, consumption and services. Naturally, the quantity of data collected

will continue to expand rapidly because of the increasing ease, availability and popularity of the web.

Data mining in the retail industry helps in identifying customer buying patterns and trends that lead to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in the retail industry –

- Design and Construction of data warehouses based on the benefits of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Analysis of the effectiveness of sales campaigns.
- Customer Retention.
- Product recommendation and cross-referencing of items.

Telecommunication Industry:

Today the telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, internet messenger, images, e-mail, web data transmission, etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business.

Data mining in the telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of the resource, and improve the quality of service. Here is the list of examples for which data mining improves telecommunication services –

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis.

Biological Data Analysis:

In recent times, we have seen tremendous growth in the field of biologics such as genomics, proteomics, functional genomics and biomedical research. Biological data mining is a very important part of Bioinformatics. Following are the aspects in which data mining contributes to biological data analysis –

- Semantic integration of heterogeneous, distributed genomic and proteomic databases.
- Alignment, indexing, similarity search and comparative analysis of multiple nucleotide sequences.

- Discovery of structural patterns and analysis of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

Other Scientific Applications:
The applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. A huge amount of data have been collected from scientific domains such as geosciences, astronomy, etc. A large amount of data sets is being generated because of the fast numerical simulations in various fields such as climate and ecosystem modelling, chemical engineering, fluid dynamics, etc.

Following are the applications of data mining in the field of Scientific Applications –

- Data Warehouses and data preprocessing.
- Graph-based mining.
- Visualization and domain-specific knowledge.