Results and Discussions

CHAPTER 4
RESULTS AND DISCUSSIONS

The demand for handling web documents has increased dramatically in recent years, owing to the rapid growth of online information used by the user. This in turn has made navigation pattern as one of the key for handling and organizing web data. The sessions identified also provide useful information. In this research work, Flame which is a sequential based algorithm is applied for web page navigation prediction. To analyze the performance of the proposed algorithm, several experiments were conducted. This chapter discusses about the data set, experimental setup which is used to analyze the data, the results obtained during performance analysis.

4.1 DATA SET

The data set used for the project is the wiki access. It contains the information about the usage of the pages by the user. They will provide the detail of the page used by the user at particular time. The data get collected from the web site WWW. Pathalizer.com.

4.1.1 Description of data set

- The attributes present in the data set are user id, page requested and the time zone.
- Except the request others are present in the numeric value.
- The user id can be obtained from the IP address which helps to separate the users.
- The time zone present in the date, year, hour, month, second format
- The get method is used to provide a request to the resource which the user needs.
- The dataset present in rows are splitted in to columns with different attributes and they are taken for the process.
- The space is used to split the various attributes in to a column.
- This help in better prediction of the users referring the pages and the link between the pages referred by them in a time.
- The navigation patterns also get predicted from the data set by analyzing the pattern of the user's page.

4.2 EXPERIMENTAL SET UP

The models which are proposed in this work are implemented by using the Matlab software.

Hardware:

Processor	:	Dual Core.
Speed	:	Above 2.30 GHz.
RAM capacity	:	2 GB.
Hard disk drive	:	500 GB.
Key Board	:	Samsung 108 keys.
Mouse	:	Logitech Optical Mouse.
Printer	:	Desk Jet HP.
Motherboard	:	Intel.
Cabinet	:	ATX.
Monitor	:	17" Samsung.

Software:

Operating system	:	Windows7 / Windows 8 & above
Coding Tool / Language	:	Matlab

4.3 MATLAB

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python. Run of the mill utilizes include:

- Math and calculation
- Algorithm improvement
- Modeling, reenactment, and prototyping
- Data investigation, investigation, and representation
- Scientific and building design

- Application improvement, including Graphical User Interface building

MATLAB is an intuitive framework whose essential information component is a cluster that does not require dimensioning. This permits you to take care of numerous specialized figuring issues, particularly those with lattice and vector definitions, in a small amount of the time it would take to compose a project in a scalar no interactive dialect, for example, C or FORTRAN.

The name MATLAB remains for framework research facility. MATLAB was initially composed to give simple access to grid programming created by the LINPACK and EISPACK ventures, which together speak to the best in class in programming for network calculation.

MATLAB has developed over a time of years with contribution from numerous clients. In college situations, it is the standard instructional device for basic and propelled courses in arithmetic, designing, and science. In industry, MATLAB is the apparatus of decision for high-efficiency examination, advancement, and investigation.

MATLAB highlights a group of use particular arrangements called tool compartments. Important to most clients of MATLAB, tool compartments permit you to learn and apply specific innovation. Tool compartments are exhaustive accumulations of MATLAB capacities (M-records) that extend the MATLAB environment to tackle specific classes of issues. Zones in which tool kits are accessible incorporate sign handling, control frameworks, neural systems, fluffy rationale, wavelets, reenactment, and numerous others.

4.3.1 MATLAB System

The MATLAB framework comprises of five principle parts:

- The MATLAB dialect.

This is an abnormal state lattice/exhibit dialect with control stream articulations, capacities; information structures, information/yield, and protest arranged programming highlights. It permits both "programming in the little" to quickly make straightforward discard projects, and "programming in the huge" to make finish vast and complex application programs.

- The MATLAB working environment.

This is the arrangement of apparatuses and offices that you work with as the MATLAB client or software engineer. It incorporates offices for dealing with the variables in your workspace and importing and trading information. It additionally incorporates devices for creating, overseeing, troubleshooting, and profiling M-records, MATLAB's applications.

- Handle Graphics.

This is the MATLAB representation framework. It incorporates abnormal state charges for two-dimensional and three-dimensional information representation, picture preparing, activity, and presentation illustrations. It additionally incorporates low-level summons that permit you to completely alter the presence of design and additionally to manufacture complete Graphical User Interfaces on your MATLAB applications.

- The MATLAB scientific capacity library.

This is an inconceivable gathering of computational calculations running from rudimentary capacities like entirety, sine, cosine, and complex number-crunching, to more modern capacities like framework opposite, grid Eigen values, Bessel capacities, and quick Fourier changes.

- The MATLAB Application Program Interface (API).

This is a library that permits you to compose C and Fortran programs that cooperate with MATLAB. It incorporate offices for calling schedules from MATLAB (dynamic connecting), calling MATLAB as a computational motor, and for perusing and composing MAT-documents.

- Language structure

The MATLAB application is worked around the MATLAB scripting dialect. Regular use of the MATLAB application includes utilizing the Command Window as an intuitive numerical shell or executing content documents containing MATLAB code.

- Variables

Variables are characterized utilizing the task administrator, =. MATLAB is a pitifully written programming dialect since sorts are certainly converted. It is a deduced wrote dialect since variables can be doled out without announcing their sort, aside from on the off chance that they are to be dealt with as typical objects, and that their sort can change. Qualities can

originate from constants, from calculation including estimations of different variables, or from the yield of a capacity.

4.4 PERFORMANCE EVALUATION

To evaluate the performance of the proposed model, six performance metrics, as listed below, are used.

- Average pattern: This is used to calculate the patterns discovered by each algorithm process. The pattern varies depending on the algorithm used, as the variation in the mining process and accuracy in the prediction helps in the computation of the average pattern prediction.

- Session Identification: The more number of related sessions helps in better identification and finding the relevant information in accurate manner with less navigation.

  Precision, recall, accuracy and F-measure are calculated from four values, namely, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) during analysis of performance.

- Precision, also known as the fraction of retrieved instances that are relevant.

$$\text{Precision(P)} : \frac{Tp}{Tp+Fp}$$

- Recall is the fraction of relevant instances that are retrieved.

$$\text{Recall (R)}: \frac{Tp}{Tp+Fn}$$

- Accuracy value is obtained from the precision and the recall value.

$$\text{Accuracy}: \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$$

- The F-measure is calculated from two measures, precision and recall.

$$\text{F-Measure}: 2\,\frac{P*R}{P+R}$$

4.5 OBSERVED RESULTS

4.5.1 Average number of patterns and sessions identified

Average number of patterns and session identified value obtained for DBscan, FP-growth and Flame algorithm are shown in Table 4.1

TABLE 4.1

Performance comparison of DBscan, FP-growth and Flame algorithm

Parameters	DB scan algorithm	FP-growth algorithm	Flame algorithm
Average Patterns	31.2445	38.6333	45.3333
Session Identified	8	10	12

The preprocessing step results in the reducing the number of url. Pattern prediction is easier while removing the unwanted data present in the log. The Figure 4.1 represent the average pattern prediction in three methods. The flame algorithm results in higher number of predicted patterns as they are supported with better suggestions for the user related to their search.

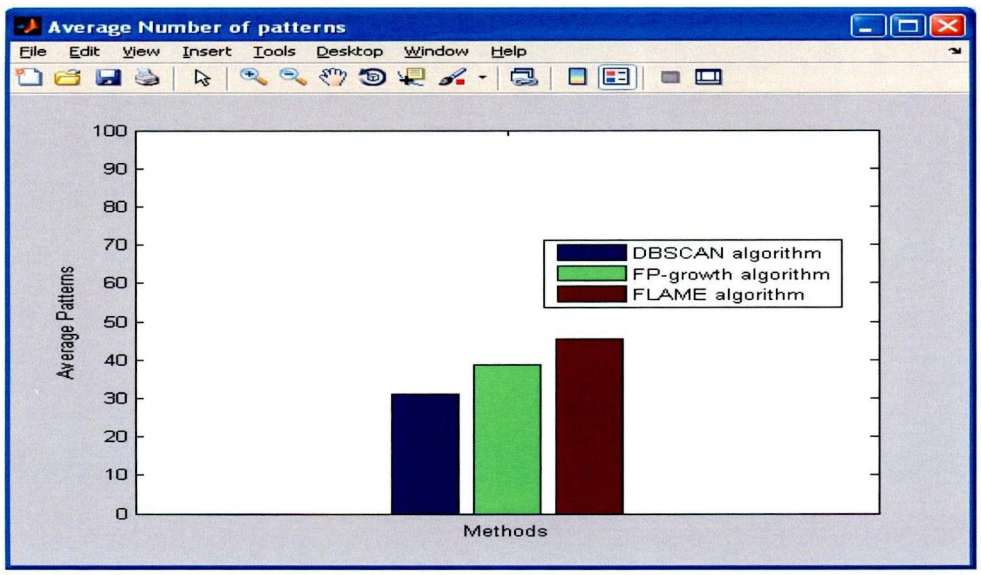

Figure 4.1 Average number of patterns comparison

The flame algorithm shows an improvement by a factor of 7 and 14 over Fp-growth and DBscan respectively in terms of average patterns identified.

The session identified by each process is shown in the Figure 4.2. The flame process shows the higher number of sessions identified. This process is used to support more number of page suggestion related to the user search by using the string match and comparison process. Figure 4.2 represent the session identified by each method.

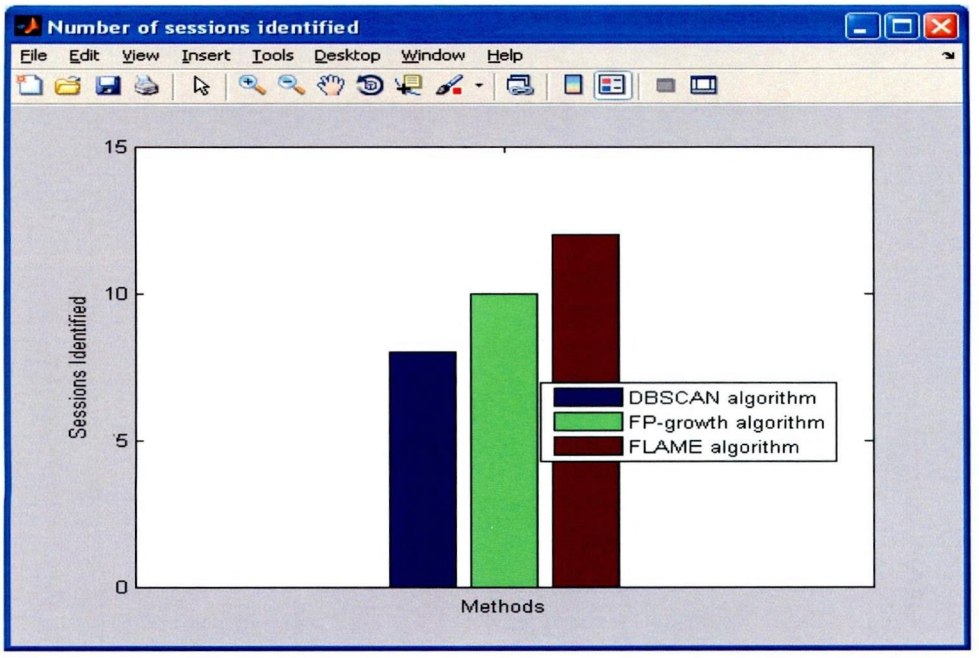

Figure 4.2 Number of Sessions identified

The flame algorithm displays an improvement by a factor of 2 and 4 over Fp-growth and DBscan respectively in terms of number of sessions identified.

4.5.2 Precision and Recall

The precision and recall value obtained for the three algorithms are shown in Table 4.2

TABLE 4.2

PRECISION AND RECALL

	DBscan	FP-Growth	Flame
Precision	76.19	83.12	90.89
Recall	80.14	93.27	97.22

From the values obtained during experimentation, though both DBscan and FP-growth produces good precision and recall. The Flame shows significant improvement by a factor of 7, 14 in precision and 4, 17 in recall over FP-growth and DBscan algorithm respectively.

Figure 4.3 shows the precision and recall of DBscan, FP-growth and Flame algorithm. From the figure, it is clear that the Flame algorithm has improved the precision and recall values.

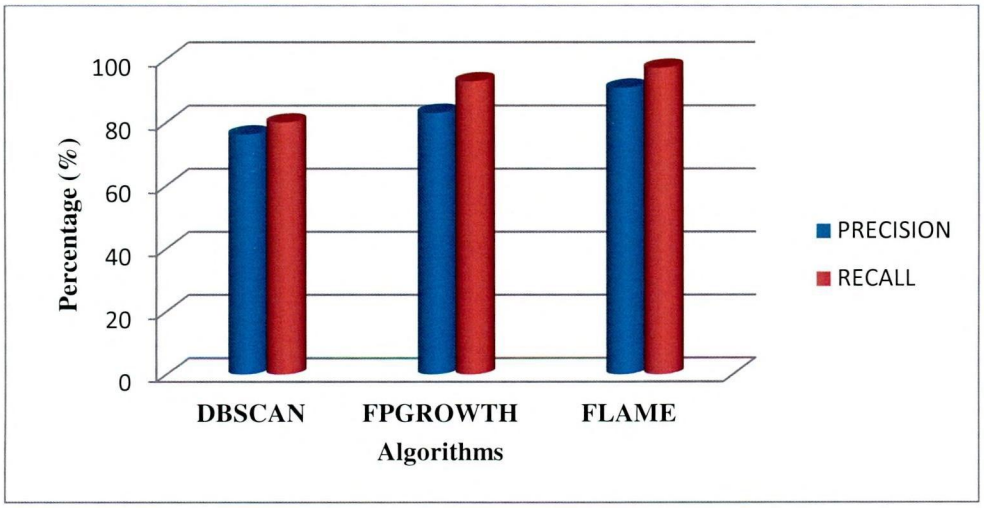

Figure 4.3 Precision and Recall metrics

4.5.3 Accuracy and F-measure

The accuracy and F-measure value obtained for three algorithms are shown in Table 4.3

Table 4.3

ACCURACY AND F-MEASURE

	DBscan	FP-Growth	Flame
Accuracy	78.25	84.24	93.75
F-measure	82.14	87.71	95.56

From the values obtained during experimentation, the Flame shows significant improvement with respect to accuracy and F-measure.

Figures 4.4 and 4.5 shows that the accuracy and F-measure of DBscan, FP-growth and Flame algorithm. From the result, it can be seen that the Flame algorithm shows an improvement by a factor of 9, 15 in accuracy and 8, 13 in F-measure over Fp-growth and DBscan algorithm respectively.

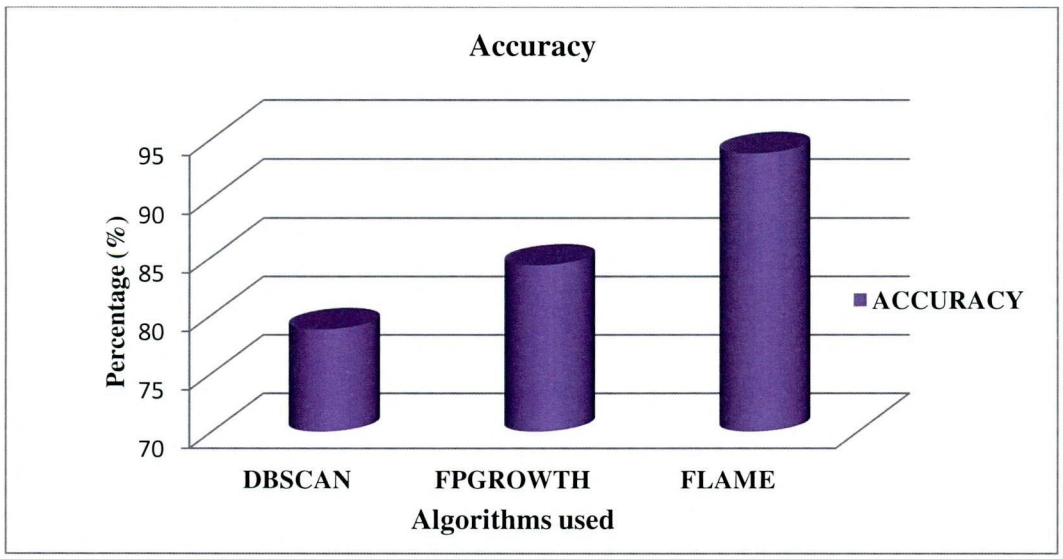

Figure 4.4 Accuracy of Patterns Predicted

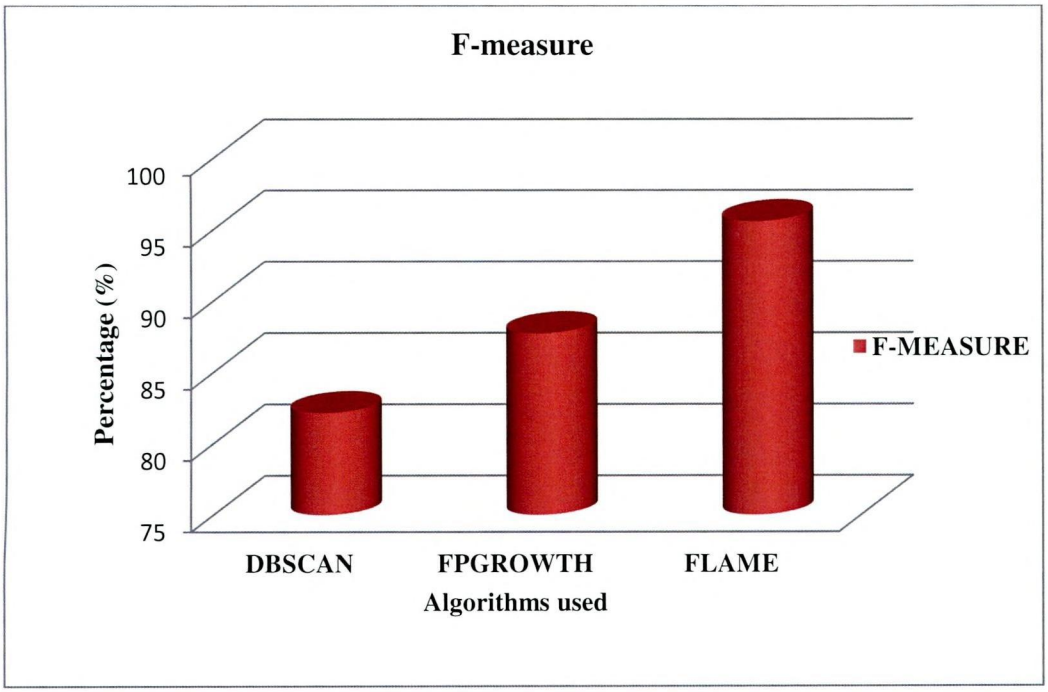

Figure 4.5 F-Measure of Patterns Predicted

4.6 CHAPTER SUMMARY

This chapter presented the experimental results of the proposed flame algorithm and analyzed various methods when applied to web log files. Five parameters namely, average number of patterns, session identified, precision, recall and F-measure were used for experiment. The results from the various experiments show that the flame algorithm used for webpage navigation pattern prediction produces significant improvement in result when compared with DBscan and FP-growth algorithm.

The next chapter concludes the work with some future research directions.