Automated Syllabus Classification in Computer Science: A Machine Learning Approach

Dr. Kavita Dhakad[1], Dr. Dipali Meher[2]
[1]*Asst. Professor, School of Information Technology, Indira University, Pune, India*
[2]*Asst. Professor, Department of Computer science, Modern College of Arts, Commerce andScience, Pune, India*

***Abstract*—This study aims to develop a comprehensive and structured repository of undergraduate Computer Science syllabi from universities, autonomous institutions, and affiliated colleges across Maharashtra. The primary objective is to systematically collect and organize these syllabi, primarily sourced from institutional websites using web crawling techniques. Leveraging advanced text mining methods, the collected syllabi are transformed into a curated dataset designed for training machine learning models. This dataset is then utilized for automated classification of Computer Science courses through state-of-the-art text classification algorithm. The outcomes of this research are expected to enhance transparency, accessibility, and consistency in Computer Science curricula across the region. Additionally, the resulting dataset will serve as a valuable resource for educators, researchers, and policymakers, promoting the integration of machine learning in educational data analysis and supporting data-driven curriculum development and evaluation. Machine learning models are designed and implemented through Python. The Best fit model is evaluated.**

***Index Terms*—text mining, text classification, web crawling, machine learning, support vector machine, support vector machine classifier**

I. INTRODUCTION

In the digital age, both individuals and organizations produce vast volumes of data daily, with an estimated 80% being unstructured. This form of data is typically disorganized, difficult to search and challenging to manage, often rendering it underutilized or entirely inactive. Consequently, organizations face significant challenges in organizing, classifying, and extracting meaningful insights from such raw information. Text mining has emerged as a powerful solution to these challenges, enabling efficient processing and analysis of unstructured data.

In business contexts, unstructured text data originates from diverse sources such as emails, social media posts, live chats, support tickets, and survey responses. Manually organizing this data is not only time-consuming and costly but also prone to human error. By contrast, text mining offers a scalable and cost-effective alternative, delivering reliable results with improved accuracy and faster processing speeds.

The dynamic nature of computer science education necessitates a comprehensive and accessible repository of undergraduate syllabi. Such a centralized resource would significantly aid curriculum development and enhance access to educational materials for both instructors and students. However, compiling and maintaining this repository manually is labour intensive and susceptible to inconsistencies. To address this issue, this study proposes the use of web crawling techniques to automate the extraction of syllabi from the official websites of educational institutions.

Following the collection phase, the unstructured syllabus documents must be transformed into a dataset suitable for machine learning applications. This involves applying text mining techniques to pre-process and convert the textual content into a structured format. The resulting dataset will support the development of robust machine learning models aimed at classifying computer science courses based on their content.

Accurate classification of computer science syllabi holds significance for multiple applications, including curriculum standardization, intelligent course recommendation systems, and academic research. This study aims to implement advanced text

classification techniques to automatically categorize course content according to predefined domains. Machine learning algorithms will be employed to develop models capable of assigning course syllabi to appropriate categories with high precision.

Machine learning-based classification systems rely on prior data specifically, training data comprising labelled examples. These examples are converted into numerical vectors, with associated tags serving as target outputs. These vectors are input into a machine learning algorithm to train a model that can subsequently identify relevant features in unseen data and predict corresponding categories. Commonly used algorithms for text classification include Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), and Deep Learning (DL) techniques.

This approach not only enhances the organization and accessibility of educational content in computer science but also contributes meaningfully to pedagogical resources and academic research.

Research Methodology

Research Questions:

1. To automate the collection of undergraduate Computer Science syllabi using web crawling
2. To build a centralized and accessible repository.
3. To transform unstructured syllabus content into a structured dataset through text mining for machine learning applications.
4. To implement a machine learning model for Automatic syllabi classification.
5. To analyse trending computer science subjects in UG Programs under study.

II. LITERATURE REVIEW

The researcher employed the Systematic Literature Review (SLR) method, which is widely recognized as a prominent research approach for investigating concepts, techniques, and algorithms in the field of text classification. SLR provides a well-structured process for identifying, evaluating, and interpreting all relevant evidence from the available literature pertaining to a specific research topic. In this study, we adhere to the SLR procedure to ensure a comprehensive and systematic analysis of the subject matter.

Search Term

TITLE-ABS-KEY (("Text Classification" OR "Text Mining") And ("Algorithm"))

Data Sources

We conducted a search query across various digital libraries, including Scopus, Web of Science, IEEE, and Google Scholar, to retrieve relevant papers for our study. These libraries serve as primary sources of literature containing potentially valuable studies on Text Mining and Text Classification. In our search, we utilized our predefined search terms targeting the title, keywords, and abstract of the papers. It is important to highlight that Google Scholar was chosen as a data source due to its inclusion of articles, research papers, and conference papers.

Data Selection

Researcher has selected 31 research papers which are covering the research objectives. The research papers cover the concepts, techniques, algorithms and application area of Text Mining and Text Classification. The research articles published since 2005 to 2022 are taken into consideration.

After examining 31 research articles, the author has gained an understanding of the multitude of text classification algorithms and application area suggested by researchers for the classification of diverse documents.

Text Classification Algorithm:

Table 1: Summary of algorithms and researcher contribution

Algorithm	#Count
Naive Bayes [NB]: [1], [2], [3], [5], [6], [7], [8], [10], [11], [13], [14], [16], [18], [19], [20], [21], [23], [24], [25], [27], [30], [31]	22
Support Vector Machines [SVM]: [1], [2], [3], [5], [6], [7], [8], [10], [11], [12], [13], [15], [16], [18], [19], [21], [22], [23], [24], [25], [27], [28], [29], [31]	24
Decision Tree [DT]: [1], [3], [5], [6], [7], [8], [11], [13], [14], [19], [23], [24], [27]	13
Random forest [RF]: [1], [2], [5], [7], [9], [10], [18], [19]	8
Deep Learning [DL]: [1], [2], [3], [4], [17], [19]	6

Author has analysed out of 31 review articles that text classification algorithms features, limitations and applications are studied. It is observed that SVM and NB are widely used for text classification of various types of documents.

Application Areas:
Author has observed the evidences of applications of above-mentioned text classification algorithms as: Syllabus classification, Job Posting Classification, News Classification, Social media documents, Health Care Domain, Research/Technical Paper and Document (General) Classification. The analysis of this is presented in Table 2.

Table 2: The types of documents where machine learning algorithms used

Application	Papers Addressed	#Count
Syllabus Classification	[4], [22], [27], [28], [29], [30]	6
Job Posting Classification	[7], [10]	2
News Classification	[9], [14], [15], [20], [21]	5
Social media documents	[12], [15],[18], [20]	4
Health Care Domain	[23]	1
Research/Technical Paper	[25], [26]	2
Document (General) Classification	[13], [15], [20]	3

III. METHODOLOGY

This research involves quantitative and qualitative research methodology for achieving the objective of the research:
Data Collection:
1. Primary Data Collection
- Syllabi Files
- Data Set in CSV Format

2. Secondary
- Articles, Books, Circulars, Newsletter, Websites etc
- Design and Implement Machine Learning based model for automatic syllabus classification by using Python libraries
- Test the Performance of the Machine Learning model by using Performance Parameter given in Python libraries.

IV. RESULT AND DISCUSSION

System Work Flow:
Step-1: Explore the Elements and Process of Syllabi Design
Step-2: Web crawling to Search syllabi on Universities/Institution websites.
Step-3: Download and compile the syllabi into a centralized repository.
Step-4: To create data set for computer science course syllabi of programs under study with essential features using text mining techniques.
Step-5: To design and implement a machine learning model for Automatic syllabus classification.
1. Multinomial Naïve Bayes (MNB)
2. Support Vector Classifier (SVC)
3. Random Forest (RF)

Step-6: To analyse the best fit Machine Learning model for syllabi classification.
Step-7: To analyse trending computer science courses in UG programs under study.

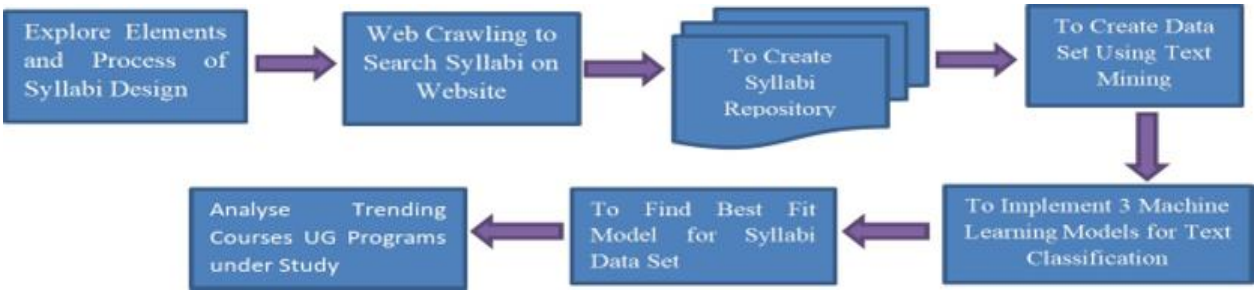


Fig 1: System Work Flow Diagram

Step-1: Explore the Elements and Process of Syllabi Design

Primary data is collected from academicians to understand elements of Good Syllabi Design by using survey method. The survey is distributed to the academicians from State Universities, Private universities and Autonomous colleges. The researcher received 125 responses due to the limited scope of the study. Secondary data collected from articles, newsletters, websites, circulars of different Government Agencies: University Grants Commission (UGC), State University, Govt. of Maharashtra etc.

Results:

Researcher analysed responses and understood the elements of good syllabus design: Name of University/Institute, Program name, Course Name, Course objectives, Learning Outcome, Total Credits, Total number of lectures, Academic year, Semester, Total Marks, Reference books and the syllabus is distributed to all the stakeholders through institute website.

The survey has enlightened that syllabi is available for all the stakeholders on University/Institution website.

Step-2: Web crawling to Search syllabi on Universities/Institution websites.

In the research context the syllabi of the various programs are publically available on Universities and Institutes websites. To collect the syllabi of programs under study from websites researcher used web crawling technique.

Automated web crawling technique is the technique involved using software to gather data from online sources automatically. The implementation is done in python with the 'Googlesearch' library. In this all the search terms are combined and used for syllabi search.

Result: Output after execution of Web Crawling Python Script.

```
Maharashtra University BSc Syllabus
https://nmu.ac.in/Student-Corner/Academics/Syllabi
https://old.nmu.ac.in/StudentCorner/Academics/Syllabi.aspx
https://mu.ac.in/syllabus
http://systelonline.org/nmu.php
https://www.shiksha.com/university/north-maharashtra-university-jalgaon-22122/courses/bsc-bc
https://ssacscjalgaon.ac.in/public/syllabus/bca.pdf
https://www.youtube.com/watch?v=o7cn0KirsG4
https://intranet.muhs.ac.in/syllabus1.aspx?c_id=150100&c_name=Basic%20B.SC.%20NURSING%20(%20Basic%20Bachelor%20of%20Science%2
0in%20Nursing)
https://www.careers360.com/university/kavayitri-bahinabai-chaudhari-north-maharashtra-university-jalgaon/courses/bsc-idpg
https://nmujalgaon.blogspot.com/2009/11/north-maharashtra-university-syllabus.html
https://www.ssvpsengg.ac.in/uploads/syllabus/1703399006_708140358.pdf
https://www.sgpcsakri.com/assets/admin/images/category_pdf/file_1652774692.pdf
https://prgscience.com/IQAC/doc2.pdf
https://sscoetjalgaon.ac.in/public/pdfs/student-corner/syllabus/2020-21_T.E_Biotechnology_Engineering.pdf
https://mitmins.edu.in/syllabus-i-first-year-basic-b-bsc-nursing/
https://sgbau.ac.in/Syllabus/syllabus.aspx
https://mvp.edu.in/kdspagri/pdfs/ay18_19/syllabus/Sullabus-B-Sc-(Hons)-5th-Dean-Agriculture--2-8-2017--Final.pdf
```

Fig 2: Output of Web Crawling Python script.

Step-3: Download and compile the syllabi into a centralized repository.

The primary data for the further research are syllabi files; these are collected as and downloaded from the websites of Universities/Institutions given by Python WebCrawler script discussed in step-2. The files were in different formats: word, pdf, images etc. The course syllabi implemented from academic year 2019-2020 to 2023-2024 are considered for study.

Result: The files are analysed and converted into the pdf formats by using online open-source tools. The researcher has created the repository "syllabus_directory" including 731 pdf files having syllabi of the courses included in program structure of programs under study. The below image represents the "syllabus_directory" repository

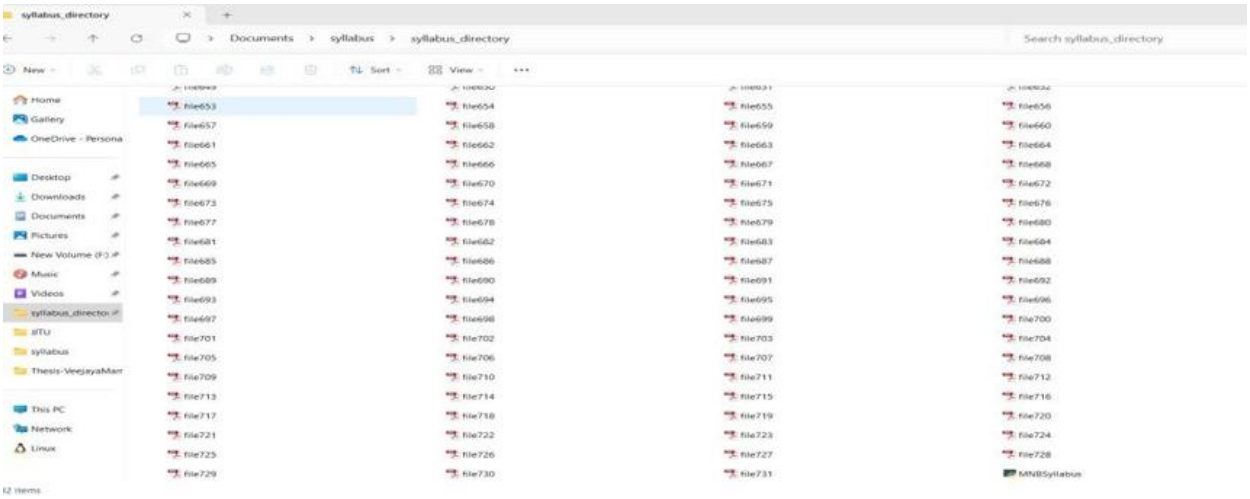


Fig 3: The File repository.

Step-4: To create data set for computer science course syllabi of programs under study with essential features using text mining techniques.

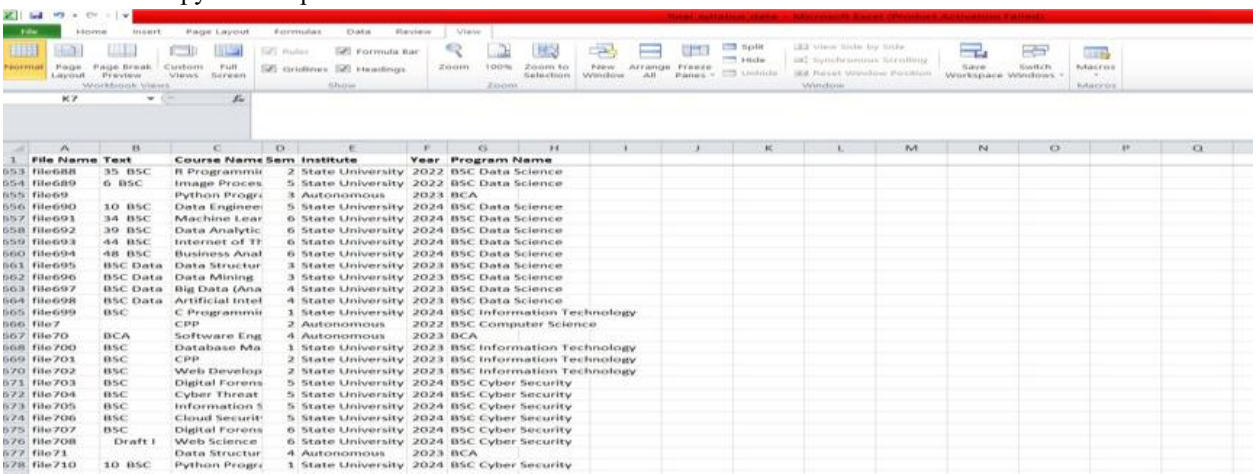
Machine Learning Model needs data set for classifying syllabi. This dataset is used for training and testing of Machine learning model. The researcher has created dataset in .csv format by using text mining techniques. The dataset is structured and ladled. The most of the features in dataset are in text form.

Result:

Table 3: Showing Feature Details

Feature Name	Feature Details	Data Type
File name	The file names mined by text mining technique which is used to store in common repository	String
Text	Content of the syllabus file mined by text mining technique	Text
Program Name	Program name mined by text mining on file content feature	String
Year	Year of the syllabus implementation is mined from the file	Integer
Course Name	The course name of the syllabus file	String
Sem	The semester for which that syllabus is created and implemented	Integer
Institute	The institute type from which the course syllabus is created	String

The Python script is executed by using text mining techniques to extract required features from course syllabi files stored in the repository and prepared data set. The image below shows content of the "final_sullabus_data.csv" after execution of the python script



	File Name	Text	Course Name	Sem	Institute	Year	Program Name
653	file688	35 BSC	R Programmir	2	State University	2022	BSC Data Science
654	file689	6 BSC	Image Proces	5	State University	2022	BSC Data Science
655	file69		Python Progr	3	Autonomous	2023	BCA
656	file690	10 BSC	Data Enginee	5	State University	2024	BSC Data Science
657	file691	34 BSC	Machine Lear	6	State University	2024	BSC Data Science
658	file692	39 BSC	Data Analytic	6	State University	2024	BSC Data Science
659	file693	44 BSC	Internet of Th	6	State University	2024	BSC Data Science
660	file694	48 BSC	Business Anal	6	State University	2024	BSC Data Science
661	file695	BSC Data	Data Structur	3	State University	2023	BSC Data Science
662	file696	BSC Data	Data Mining	3	State University	2023	BSC Data Science
663	file697	BSC Data	Big Data (Ana	4	State University	2023	BSC Data Science
664	file698	BSC Data	Artificial Intel	4	State University	2023	BSC Data Science
665	file699	BSC	C Programmir	1	State University	2024	BSC Information Technology
666	file7		CPP	2	Autonomous	2022	BSC Computer Science
667	file70	BCA	Software Eng	4	Autonomous	2023	BCA
668	file700	BSC	Database Ma	1	State University	2023	BSC Information Technology
669	file701	BSC	CPP	2	State University	2023	BSC Information Technology
670	file702	BSC	Web Develop	2	State University	2023	BSC Information Technology
671	file703	BSC	Digital Forens	5	State University	2024	BSC Cyber Security
672	file704	BSC	Cyber Threat	5	State University	2024	BSC Cyber Security
673	file705	BSC	Information S	5	State University	2024	BSC Cyber Security
674	file706	BSC	Cloud Securit	5	State University	2024	BSC Cyber Security
675	file707	BSC	Digital Forens	6	State University	2024	BSC Cyber Security
676	file708	Draft I	Web Science	6	State University	2024	BSC Cyber Security
677	file71		Data Structur	4	Autonomous	2023	BCA
678	file710	10 BSC	Python Progr	1	State University	2024	BSC Cyber Security

Fig 4: Automatically extracted Data Set

Step-5: To design and implement a machine learning model for Automatic syllabus classification.

Machine learning algorithms extensively used for text classification encompasses Naive Bayes (NB), Support Vector Classifier (SVC), Decision Trees (DT), Random Forests (RF), and Deep Learning (DL) techniques. The selection of algorithm depends on various factors like data set size, features etc.

In this research literature review, the author has rigorously reviewed application areas where above listed algorithms have been used successfully, such as syllabus classification, job posting classification, news classification, social media documents, the healthcare domain, research/technical papers, and general document classification. Notably, the findings highlight that SVC, NB and RF has performed well particularly for text classification. Following image shows the steps used to prepare Machine learning Model.

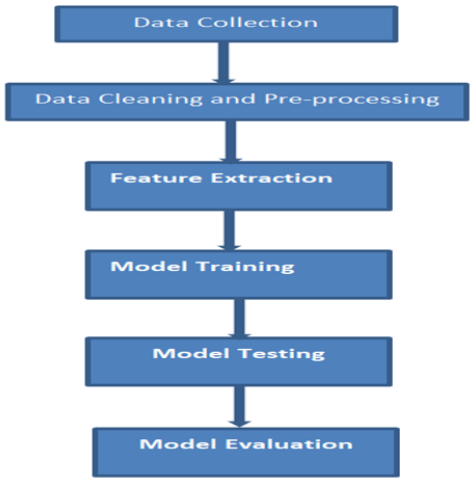


Fig 5: Machine Learning Model Design flow

Results:

The researcher has prepared ML Model with selected Text Classification Algorithms by adopting concepts as: Multinomial/Multiclass Classification, Feature Extraction, Cross-Validation, Divide the Dataset into Training and Testing Datasets The implementation is done with the Python Programming Language libraries.

Various ML models can be used for text classification, including but not limited to Logistic Regression, Navie Bayes (NB), Support Vector Classifier (SVC), Random Forests (RF) and Neural Networks. The choice of model depends on factors such as the size and complexity of the dataset. The literature review concludes that SVM is working good for the text classification. The researcher has designed and implemented ML Model by using Multinomial NB, SVC and RF algorithms. Machine Learning Model implementation is done by using Python libraries.

Step-6: To analyse the best fit Machine Learning model for syllabi classification.

The researcher has designed and implemented ML Model by using Multinomial NB, SVC and RF algorithms. These models are trained and tested during the process. The best model is selected based on its accuracy and Confusion matrix evaluated.

Results:

1. Multinomial NB Algorithm Confusion Matrix and Accuracy

	precision	recall	f1-score	support
BCA	0.61	0.93	0.74	55
BSC Computer Science	0.67	0.81	0.74	43
BSC Cyber Security	0.00	0.00	0.00	3
BSC Information Technology	0.00	0.00	0.00	13
BSC Data Science	1.00	0.05	0.09	22
accuracy			0.64	136
macro avg	0.46	0.36	0.31	136
weighted avg	0.62	0.64	0.55	136

2. Support Vector Classifier Algorithm Confusion Matrix and Accuracy

```
Average accuracy for Naive Bayes Text Classification: 63.68
```

3. Random Forest Algorithm Confusion Matrix and Accuracy

	precision	recall	f1-score	support
BCA	1.00	0.95	0.97	58
BSC Computer Science	0.90	1.00	0.95	44
BSC Cyber Security	1.00	1.00	1.00	8
BSC Information Technology	1.00	0.50	0.67	4
BSC Data Science	0.86	0.86	0.86	22
accuracy			0.94	136
macro avg	0.95	0.86	0.89	136
weighted avg	0.94	0.94	0.94	136

```
Average accuracy for Random Forest Text Classification: 94.37
```

Step-7: To analyse trending computer science courses in UG programs under study.

The dataset and the results of the model are analyzed. Researcher is interested in finding the current trending program offered by Universities/Institutes under study. Further the researcher is intending to find coursed offered in these trending programs. The python script is written to find the frequency count by year and program wise.

Results:

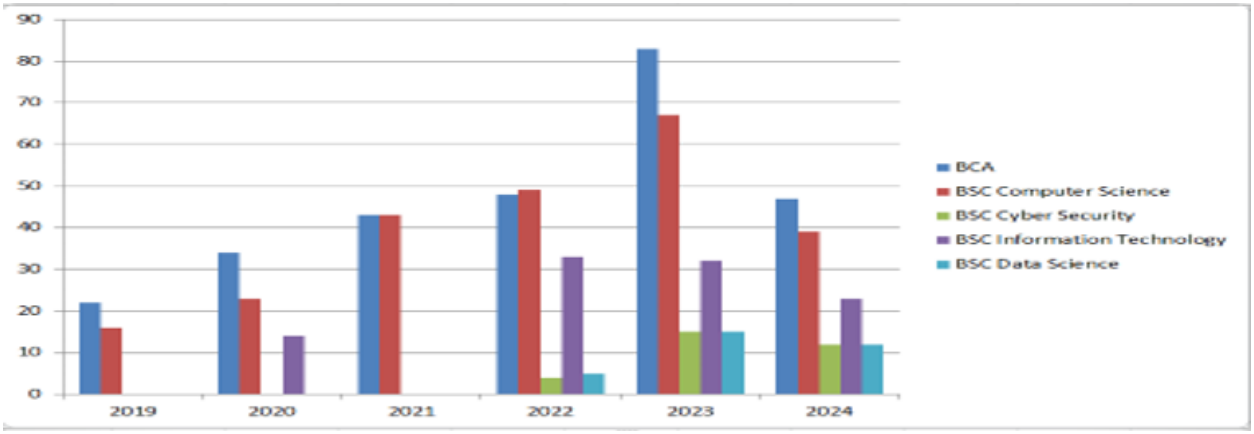


Fig 6: The graph shows the Data Visualization year wise frequency count of the Program

BSC Data Science and BSC Cyber Security has introduced in 2022 it is gradually growing in 2023 and 2024. The NEP 2020 is introduced by all of the universities in Maharashtra state, so program structure has changed in 2024. The subject in these trending programs is introduced in 2023 and having frequency count one. It clearly observed that these are unique and latest subjects.

V. CONCLUSION

There is need of automation in the academic's field for syllabi designing process. Researcher has given automation to some of the steps: Searching by web crawling, data collection and classification. The researcher collected 731 syllabi PDFs which were successfully downloaded using web scraping techniques, forming the initial dataset for this study. After an extensive data cleaning process, text mining was applied to these files, resulting in a refined dataset

of 678 records. The dataset is created 7 features which will be further used for Machine Learning Model.

The Machine Learning Models has designed and implemented for Program classification by using Multinomial Navie Bayes, Support Vector Classifier and Random Forest algorithms.

The performance is average of 20 iterations by using cross validation techniques. Multinomial Naive Bayes (MNB) exhibited overall accuracy of 63.67%. Support Vector Classifier (SVC) exhibited overall accuracy of 72.72%. The model demonstrated good performance. Random Forest (RF) revealed overall accuracy of 94.37%. This high level of accuracy underscores the potential of automated techniques in handling large volumes of educational data, thereby streamlining the process of syllabus classification and enhancing the organization of educational resources.

Future Scope:

Future work can involve expanding the dataset by including syllabi from a wider range of disciplines beyond computer science. Deep learning techniques can be used to improve the accuracy of the model.

REFERENCES

[1] Li, Q., H. Peng, J. Li, C. Xia, R. Yang, L. Sun, et al., "A survey on text classification: From traditional to deep learning," ACM Transactions on Intelligent Systems and Technology, vol. 13, no. 2, pp. 1–41, 2022.

[2] Sarker, I. H., "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science, vol. 2, no. 3, Art. no. 160, 2021.

[3] Bhavani, A., and B. S. Kumar, "A review of state-of-the-art text classification algorithms," in Proc. 5th Int. Conf. Computing Methodologies and Communication (ICCMC), 2021, pp. 1484–1490.

[4] Yasukawa, M., H. Yokouchi, and K. Yamazaki, "Syllabus mining for analysis of searchable information," International Journal of Institutional Research and Management, vol. 4, no. 1, pp. 46–65, 2020.

[5] Maleki, F., K. Ovens, K. Najafian, B. Forghani, C. Reinhold, and R. Forghani, "Overview of machine learning—Part 1: Fundamentals and classic approaches," Neuroimaging Clinics, vol. 30, no. 4, pp. e17–e32, 2020.

[6] Gupta, R., "A survey on machine learning approaches and its techniques," in Proc. IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS), 2020, pp. 1–6.

[7] Nasser, I., and A. H. Alzaanin, "Machine learning and job posting classification: A comparative study," International Journal of Engineering and Information Systems, pp. 6–14, 2020.

[8] Mahesh, B., "Machine learning algorithms—A review," International Journal of Science and Research, vol. 9, pp. 381–386, 2020.

[9] Shah, K., H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for text classification," Augmented Human Research, vol. 5, Art. no. 12, 2020.

[10] Mittal, S., S. Gupta, A. Shamma, I. Sahni, and D. N. Thakur, "A performance comparison of machine learning classification techniques for job titles using job descriptions," 2020.

[11] Ray, S., "A quick review of machine learning algorithms," in Proc. Int. Conf. Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 35–39.

[12] N. V. M., and D. A. Kumar R., "Implementation of text classification using bag of words model," in Proc. 2nd Int. Conf. Emerging Trends in Science and Technologies for Engineering Systems (ICETSE), 2019.

[13] Naresh, E., K. B. Vijaya, V. S. Pruthvi, K. Anusha, and V. Akshatha, "Survey on classification and summarization of documents," International Journal of Research in Advent Technology, vol. 7, no. 6S, 2019.

[14] Ababneh, J., "Application of Naïve Bayes, decision tree, and K-nearest neighbors for automated text classification," Modern Applied Science, vol. 13, no. 11, p. 31, 2019.

[15] Chatterjee, S., P. G. Jose, and D. Datta, "Text classification using SVM enhanced by multithreading and CUDA," International Journal of Modern Education and Computer Science, vol. 12, no. 1, pp. 11–20, 2019.

[16] Kadhim, A. I., "Survey on supervised machine learning techniques for automatic text

classification," Artificial Intelligence Review, vol. 52, no. 1, pp. 273–292, 2019.

[17] Luan, Y., and S. Lin, "Research on text classification based on CNN and LSTM," in Proc. IEEE Int. Conf. Artificial Intelligence and Computer Applications (ICAICA), 2019, pp. 352–355.

[18] Hartmann, J., J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," International Journal of Research in Marketing, vol. 36, no. 1, pp. 20–38, 2019.

[19] Kowsari, K., K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," Information, vol. 10, no. 4, Art. no. 150, 2019.

[20] Panda, M., "Developing an efficient text pre-processing method with sparse generative Naïve Bayes for text mining," International Journal of Modern Education and Computer Science, vol. 11, no. 9, pp. 11–19, 2018.

[21] Miao, F., P. Zhang, L. Jin, and H. Wu, "Chinese news text classification based on machine learning algorithm," in Proc. 10th Int. Conf. Intelligent Human-Machine Systems and Cybernetics (IHMSC), vol. 2, 2018, pp. 48–51.

[22] Orellana, G., M. Orellana, V. Saquicela, F. Baculima, and N. Piedra, "A text mining methodology to discover syllabi similarities among higher education institutions," in Proc. Int. Conf. Information Systems and Computer Science (INCISCOS), 2018, pp. 261–268.

[23] Allahyari, M., S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "A brief survey of text mining: Classification, clustering and extraction techniques," arXiv preprint arXiv:1707.02919, 2017.

[24] Desai, A., "A review on knowledge discovery using text classification techniques in text mining," International Journal of Computer Applications, vol. 111, no. 6, 2015.

[25] Adeva, J. G., J. P. Atxa, M. U. Carrillo, and E. A. Zengotitabengoa, "Automatic text classification to support systematic reviews in medicine," Expert Systems with Applications, vol. 41, no. 4, pp. 1498–1508, 2014.

[26] Nguyen, T. H., and K. Shirai, "Text classification of technical papers based on text segmentation," in Natural Language Processing and Information Systems (NLDB 2013), Berlin, Germany: Springer, 2013, pp. 278–284.

[27] Rathod, N., and L. N. Cassel, "Machine learning in building a collection of computer science course syllabi," in Theory and Practice of Digital Libraries (TPDL 2012), Berlin, Germany: Springer, 2012, pp. 357–362.

[28] Ota, S., and H. Mima, "Machine learning-based syllabus classification toward automatic organization of issue-oriented interdisciplinary curricula," Procedia - Social and Behavioral Sciences, vol. 27, pp. 241–247, 2011.

[29] Yu, X., M. Tungare, W. Yuan, Y. Yuan, M. Pérez-Quiñones, and E. A. Fox, "Automatic syllabus classification using support vector machines," in Handbook of Research on Text and Web Mining Technologies, Hershey, PA: IGI Global, 2009, pp. 61–74.

[30] Joorabchi, A., and A. E. Mahdi, "An automated syllabus digital library system for higher education in Ireland," The Electronic Library, vol. 27, no. 4, pp. 640–658, 2009.

[31] Ikonomakis, M., S. Kotsiantis, and V. Tampakas, "Text classification using machine learning techniques," WSEAS Transactions on Computers, vol. 4, no. 8, pp. 966–974, 2005.

[32] Rathod, N., and L. N. Cassel, "Machine learning in building a collection of computer science course syllabi," in Theory and Practice of Digital Libraries (TPDL 2012), Berlin, Germany: Springer, 2012, pp. 357–362.

[33] Scikit-learn Documentation, Version 0.21. Available: https://scikit-learn.org/0.21/documentation.html. Accessed: May 30, 2024.

[34] Natural Language Toolkit (NLTK) API Documentation. Available: https://www.nltk.org/api/nltk.html. Accessed: May 28, 2024.

[35] Bishop, C. M., Pattern Recognition and Machine Learning. New York, NY: Springer, 2006.

[36] Ramos, J., "Using TF-IDF to determine word relevance in document queries," in Proc. First Instructional Conf. Machine Learning, 2003, pp. 133–142.

[37] Kuhn, M., and K. Johnson, Applied Predictive Modeling, 2nd ed. Cham, Switzerland: Springer, 2023.