

# Text Mining Unstructured Data

Academic Counselling Solution



Enabling better  
and comprehensive  
academic counselling  
with Data Analytics

# Executive Summary

## Global Education! Bigger Opportunities!

International education broadens the student's perspectives by imparting diverse cultural skills, intercultural competencies, global competence, in addition to opening a new world of opportunities, and taking students to better career opportunities. To realize their dream of international education, students seek information about educational institutions from various sources without awareness on the consistency of available data.

However, international education is a highly confusing, complicated, involves acute competition, peer pressure, stress, and difficulty of choosing the right career choices. Quite often, this causes confusion, doubts, and negative thoughts for the students.



Education counselling is a comprehensive support system that provides unbiased, impartial, professional, and highly personalized advice on international higher education institutions and course options based on the students' academic qualifications, work experience, financial capabilities, career goals, preferred place of study, and other preferences. Education counsellors play a pivotal role in enabling students realize their international education aspirations by simplifying / demystifying international education and eliminating the complexities involved.

Students and almost all academic counsellors generally use websites, forums, and other social media sites for obtaining information and sharing their feedback

in the form of text, video, audio and so on, that are often "unstructured" in nature. As the academic institutions are expanding their academic programs and resources, the amount of data generated is equally increasing in proportion with time. They need to use the right academic counselling solution that will help them filter the required data, analyze it, derive insights, and make the right decisions.

HTC's Academic Counselling Solution creates international educational pathways to students by simplifying the task of academic counsellors. Our solution enables education counsellors to provide personalized and intelligent international education counselling with real time interactive insights. This will facilitate the students to enroll in top notch global universities along with information about scholarships / financial aid.

## Data Mining

### Simplify identifying Trends and Patterns

Data Mining is the process of using business analytics methodology to find the hidden patterns over large amount of data to achieve business goals. It also facilitates the usage of tools for automated learning (such as using Machine Learning Technique) based on historical data and patterns that helps in future decision making. There are various software available for analyzing data, which will allow the user to analyze data from various angles, classify, and summarize it effectively.



Data pre-processing includes identification of the required data, loading them from various sources and integrating, normalizing to a standard format that facilitates easier pattern recognition and analysis and cleansing of data to suit the business needs, followed by carrying out exploratory data analysis. Results could be delivered with the help of model application such as dashboards or ad hoc applications by summarizing the results of the analysis in a simple form for business users.

Data mining has greater potential in helping universities / academic institutions and students take proactive, knowledge-driven decisions. Data mining tools help them to answer questions promptly that were traditionally too time consuming.

## Educational Data Mining

### Game changer for Education Sector

When the details of multiple institutions have to be collected for decision-making, manually searching for them is a time consuming and cumbersome process. Sometimes the required information can be stored in static files, which were earlier collected for reference purposes. As this cannot be used as first-hand reference to derive information, the requirement for mining the raw data, analyzing it, so that it can be converted to meaningful information in real time is high. To address this issue, Educational Data Mining (EDM), an emerging area of study is focusing on developing methods for exploring meaningful information over the unique and increasingly large scale data that come from educational sources. EDM develops methods and applies the techniques from statistics, machine learning, and data mining to analyze data collected during teaching and learning. Text mining methodology is used to understand the exact requirements of students and parents and provide the appropriate results they are looking for to make the decision.



# Need for Big Data Analytics on Text Data

## Talking about Analytics

Big Data Analytics refers to the process of collecting, organizing, and analyzing large data sets. Institutions are focusing on identifying the methods to make effective learning, identify the learners' performance, and to predict future performance. The public's interest on transparency is necessitating educational institutions to implement big data analytics. Analytics uses data mining, pattern matching, text mining, Natural Language Processing (NLP) often with Machine learning methodologies. There is a wide range of data for academic institutions spread across the Internet. People look for customized information for prompt decision-making, which is rising interest in text data mining over websites of academic institutions across the globe.

While referring information related to academic institutions from social media or the institutional websites, most of the data is in unstructured format. They are in the form of XML, HTML, PHP, or text which provides an opportunity for business to use text mining methodologies to extract those data and to apply pattern matching mechanisms and methodologies to locate the needed data among large sets of unstructured data.



Many of the academic institutions are posting new information or recent changes in their websites as well as social media. Such data is growing at double the rate of traditional data and there is a need to keep the business updated with latest data.

# Why Text Data Mining Approach?

## Increase Potential

Text data mining will increase the progress and accuracy of analytics systems. Carrying out text mining as a part of data mining process can act as a potential for facilitating the knowledge discovery on large set of unstructured data. With growing amount of Academic disciplines and institutions, the analytics will enable us to fully realize the value of data.

## Data: 80-20 Game

### Derive Insights from Data

There is continual growth of data in the education sector. Nowadays, most educational institutions store terabytes to petabytes of data. This huge volume of data is highly complex as people tend to store data both in structured and unstructured format. The unstructured data is stored as PDF files, images, videos, and geo-spatial data. Counsellors, students, tutors, research developers, and all others who deal with educational data are challenged by the speed and format of data. They need to update data in real time to provide the right service to their customers and most of the data is in unstructured format.

The conclusion from various surveys prove that only 20% of data is in structured format, whereas, the remaining 80% of all the data from various sources are available in unstructured form. If only 20% of structured data is utilized to gain insights, then there are massive opportunities awaiting to be leveraged in the analysis of those 80% of unstructured data.

## Real time Challenges

### Unlocking the unstructured data

When data can be counted, it can be analyzed. When it can be analyzed, it can be interpreted. Unlocking this huge unstructured data represents the next Big Data challenge. Use of Analytics in real world applications is growing faster as people have started recognizing the potential of big data and analytics. The value of applying analytics on data for decision making is growing exponentially. The

unstructured data must be unlocked to effectively utilize it for deriving insights. For Text mining and NLP, Text Analytics enables converting unstructured text into structured data. It has moved from university research into real-world products that can be used by any business that is more concerned with educational data.

## **Harnessing the power of R**

### **Power of Pattern Matching**

While carrying out text mining, there is a need for proper validation of educational data that is taken from students' use of interactive learning environments, collaborative learning, reference websites or administrative data from schools and universities. There is an existence of multiple levels of meaningful hierarchy over each set of data such as admission details, fee structure, and eligibility criteria etc., which often need to be determined by properties in the data itself.

The focus must be on the quality of the text, validation of abbreviations and acronyms, validation of usage of spoken word by means of NLP, evaluating richness of content, content (with respect to some benchmark text that needs to be extracted), and similarity of content (among all other similar usage of text by using a data dictionary). For Classification, a large set of unstructured data is classified institution wise and clustering is performed to group the entities within them. To identify the relationship between the filtered out data, association rules are used to map them under different clusters. Then the data is validated by means of Sequential pattern mining with a list of keywords.

## **What makes the Difference?**

### **Analytics based Decision-making**

A few businesses are using off the shelf services to collect, analyze, and visualize the unstructured data, like with web search results or external reviews or data from advertisements. There are hundreds of businesses serving the education market and launching ad-hoc products every day. On the other hand, our analysts have an affinity for a particular business intelligence tool, many of which play nicely with text mining, pattern matching, and analytics allowing the analysts to handle unstructured data to get meaningful insights.

# Issues faced by Academic Counsellors

## Locate the Right Information

Globalization out of modern communications has major impact in the university curriculum, classrooms, and academic advising delivery systems. Academic advising is the exchange of information that helps students achieve their educational and career goals. For many students selecting a course to determine their career path is difficult and confusing. The academic counsellors are the advisors who assist students to improve their academic and career plan and discuss with them how the advised course will fit their academic interest. They also assist students in getting the required support from offices of the selected University.



The students' academic future depends on counsellor's advice, it can be a challenge to organize the various demands. Students expect to know various institution's academic programs, placements, fee structure, eligibility criteria, policies and procedures, exploring different majors, about dropping or adding a course. Currently this is carried out manually by searching each and every institution on the web. By this, many of the institutions or programs may be missed out due to time constraints or limitations. Mining a catalog manually is a difficult task. The usage of traditional approach in this technology driven market is a challenge to counsellors when they have to reach diverse set of students. Growth in technologies are changing the traditional methods in which colleges

and universities operate and so does advising services too. Using an efficient academic advising system will enhance the quality of services offered with real-time information at anytime from anywhere.

## Enabling Availability and Consistency through Academic Counselling Solution

### How are we addressing?

Students seek Academic Counsellors' support to choose the right set of universities for achieving their academic and career goals. Since every student's requirement is unique, the counsellors have to scrutinize, analyze, and direct the students carefully. The counsellors must have first-hand experience with the entire international education process. They must demystify the complexities of international education and guide the students throughout the entire course of action - selecting the right study program at the right university, academic and other related costs, visa process, and all their apprehensions about international education that matches ambitions, career plans, capabilities, eligibility criteria, and budgetary constraints.



As all the verticals are significantly important, all major companies have made clear moves showing the importance of text analytics on various fields that

includes educational products as well (for example: Microsoft's Encarta Kids software). Academic Analytics is an essential component in the existing market. Applying analytics will lead the institutions to be more transparent and intelligent with evidence.

HTC Global Services developed a web-based business intelligence "Academic Counselling Solution" that puts the expected results on the lime light and all those complex analytics are silently done behind the screen. The solution harnesses the power of text analytics on unstructured data of academic institutions situated globally. Our business objective on analyzing academic data is the origin of our data mining solution. In addition to this we are utilizing our business knowledge as a central for every step of our data mining process. We are doing our data preparation before data manipulation begins, because this process is more than half of every data mining process. Academic Counselling Solution facilitates the decision making of academic counsellors as well as the students to perform the customized search in their field of interest that fetches accurate results to their expectations.

## References

1. Global Education  
[https://en.wikipedia.org/wiki/Global\\_education](https://en.wikipedia.org/wiki/Global_education)
2. Unstructured Data  
[https://en.wikipedia.org/wiki/Unstructured\\_data](https://en.wikipedia.org/wiki/Unstructured_data)
3. Farhad Soleimanian Gharehchopogh, Zeinab Abbasi Khalifehlou, "Analysis and evaluation of unstructured data: Text mining versus natural language processing", Application of Information and Communication Technologies (AICT), 2011 5th International Conference, DOI  
[https://www.researchgate.net/publication/224266411\\_Analysis\\_and\\_evaluation\\_of\\_unstructured\\_data\\_Text\\_mining\\_versus\\_natural\\_language\\_processing](https://www.researchgate.net/publication/224266411_Analysis_and_evaluation_of_unstructured_data_Text_mining_versus_natural_language_processing)
4. Data Mining & Predictive Analytics  
[http://khabaza.codimension.net/index\\_files/9laws.htm](http://khabaza.codimension.net/index_files/9laws.htm)
5. Data mining  
[https://en.wikipedia.org/wiki/Data\\_mining](https://en.wikipedia.org/wiki/Data_mining)
6. Data mining  
<http://hanj.cs.illinois.edu/pdf/ency99.pdf>
7. Jayanthi Ranjan, Kamna Malik, (2007) "Effective educational process: a data-mining approach", VINE, Vol. 37 Iss: 4, pp.502 – 515  
<http://dx.doi.org/10.1108/03055720710838551>
8. Big Data on Campus  
[http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html?\\_r=0](http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html?_r=0)
9. A Survey of Text Mining Techniques and Applications  
<http://www.jetwi.us/uploadfile/2014/1230/20141230112729939.pdf>

10. David Alfred Ostrowski, A Framework for the Classification of Unstructured Data, Berkeley, CA, DOI : 10.1109/ICSC.2009.48, 14-16 Sept. 2009 pp. 373 – 377

<https://support.sas.com/resources/papers/proceedings14/1288-2014.pdf>

11. Big data analytics

[http://www.webopedia.com/TERM/B/big\\_data\\_analytics.html](http://www.webopedia.com/TERM/B/big_data_analytics.html)



*Reaching out... through IT*®

### **World Headquarters**

3270 West Big Beaver Road

Troy, MI 48084, U.S.A

Phone: 248.786.2500

Fax: 248.786.2515

[www.htcinc.com](http://www.htcinc.com)