PREDICTIVE ATTENDANCE ANALYTICS: A CASE STUDY

Kauleshwar Prasad [1], Arpana Rawal [2], Ani Thomas [3]

[1]*Assistant Professor, Computer Science and Engineering*
[2]*Professor, Computer Science and Engineering*
[3]*Professor, Computer Science and Engineering*

[1] kauleshwarprasad@gmail.com, [2] arpana_rawal@rediffmail.com, [3] ani.thomas@bitdurg.ac.in

Abstract

The hastily expanding field of big data analytics has been playing an important role in the field of academia. It has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured data produced by education. Educational data analytics and Learning analytics are the growing fields of study through which learning and teaching practices can be improved. Big data analytics can be used to improve education delivery and provide decision support to higher management to streamline the activities of academic semesters within deadlines. This paper focuses on assimilating highlights on introduction of big data analytics and growing area of visual analytics. Various research scopes in field of academia are assembled. A case study for analyzing the learning environment to track the attendance patterns is taken. After anticipating semester attendance patterns of each student, decision - making system is developed so that proper counseling can be done for frequent absenteeism cases.

Keywords- *Educational data analytics, Attendance analytics, data visualization*

1. INTRODUCTION

In the past two decades, data are emerging continuously both in structured and unstructured form from the various domains like health care, scientific sensors, user – generated data, Internet and financial companies. Due to explosion of these data sets, the concept of big data was coined. The emergence of big data and its paradigm has transformed our society. Big data platform is divided into four components, i.e., data generation, data acquisition, data storage, and data analysis. Data generation basically concern with the different sources through data can be generated. After the data has been collected, it should be pre - processed. Data pre – processing is done for efficient storage and mining. Data storage concerns with the storage and managing large - scale datasets. Various interface functions are provided by the storage system for the data analysis. Various analytical methods and tools are used to uncover the hidden facts. According to the processing time requirement, big data analytics can be categorized into two forms:

- *Streaming Processing:* In case of streaming processing, data arrives continuously and small portion of the stream is stored. Analysis of data is done as soon as possible to derive its results.
- *Batch Processing:* Here the data are first stored and then analyzed. MapReduce is used to process the batch data in which data are first divided into small chunks. These chunks are processed in parallel to generate intermediate results. All these intermediate results are aggregated to get the final result [1].

Visual analytics is a new research field which is a combination of both data analysis and information visualization. Visual analytics techniques combine the concepts from data mining, machine learning, human computing interactions, and human cognition.

2. EDUCATIONAL DATA ANALYTICS ON MOVE

Various researches have been done in the field of data analytics in academia. Various tools available for data analytics and data visualization are weka, data wrangler, R, Google Fusion table, Impure, Tableau Public, Protovis, Quantum GIS (QGIS) and openstructMap (LO‟AI A. TAWALBEH et al., 2016). It is helpful in analyzing the underlying patterns to predict students‟ outcome, needing extra help or need assignments for further improvement (Darrel M. West, 2012). Machine learning with data analytics can be applied on one of the massive datasets, Massive Open Online Course (MOOC) (Jiajun Liang et al., 2016). In order to improve the teaching learning process, Learning Management System (LMS) can be developed. Data can be collected from the academic institutions and prediction is done in the performance of students. Grade Performance system (GPS) can be developed in which students receive alerts regarding grades, attendance, academic issues as well as positive feedback (Anthony G. Picciano, 2012). Various methods used for the prediction are Regression based methods, Matrix factorization (RF) method, Factorization method (FM), Personalized linear multi – regression (PLMR), Random Forest (RF), Mean of means and Uniform random guessing methods (Asmaa Elbadrawy et al., 2016). Effort has also been made for developing Virtual Learning Environmens (VLEs). It is based on online teaching and learning process for which SmartLAK architecture is used (Thomas Rabelo et al. 2015).

3. PROBLEM FORMULATION

Unlike the various survey done on big data solutions upon structured and / or unstructured data streams in academia, a major research challenge arises is to evaluate the feasibility study of these solution upon scalable data sets. A case study of analyzing the learning environment to track the attendance pattern of students can be taken. After tracking the attendance pattern of student, decision - making system can be developed so that proper counseling can be done for extreme absenteeism cases.

3.1 Experimental Setup

In order to track the attendance pattern of corresponding students, attendance values would be like 1’s for presence and 0’s for absence in four dimension i.e subject, batch, month and semester. Figure 1 illustrates a snapshot of attendance database sheets of one batch of third semester DELD subject of September month. Similarly attendance of all subjects is recorded in the form of 1’s and 0’s. If the lecture is not held at particular date then don’t care (D) is entered.

		3rd Sem Digital Electronics September 2014																													
CRN	NAME OF CANDIDATE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
A - 01	VARTIKA NAYAK	1	D	1	D	1	1	D	1	D	1	D	2	0	D	1	D	0	D	0	D	D	0	D	D	D	D	D	D	0	D	D
A - 02	AKASH AGGARWAL	1	D	0	D	1	1	D	0	D	1	D	2	1	D	1	D	0	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 03	KUMARI PRIYA	1	D	1	D	1	1	D	1	D	1	D	2	0	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	0	D	D
A - 04	VIVEK KUMAR ROCKY	1	D	0	D	1	1	D	1	D	1	D	2	1	D	1	D	1	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 05	SHIKHA KUMARI JHA	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 06	AKSHAY BELE	0	D	1	D	1	0	D	0	D	1	D	2	0	D	0	D	0	D	0	D	D	0	D	D	D	D	D	D	0	D	D
A - 07	ANJALI DHRUW	0	D	0	D	0	1	D	1	D	1	D	2	1	D	1	D	0	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 08	ANURAG KUMAR SINGH	1	D	0	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 09	G. NIVEDHA NADAR	1	D	1	D	1	0	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 10	AVINASH MISHRA	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 11	CHINMAY JAIN	1	D	1	D	0	1	D	1	D	0	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 12	DEEP SINGH	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	0	D	D
A - 13	DEEPSHIKHA KURRE	1	D	1	D	1	1	D	1	D	1	D	2	0	D	0	D	1	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 14	DEVIKA	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 15	DIKESH KUMAR	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	0	D	D	D	D	D	D	0	D	D
A - 16	DUSHAYANT KUMAR KOTHARI	1	D	1	D	1	1	D	1	D	1	D	2	0	D	0	D	1	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 17	HANSHIKA AGRAWAL	1	D	1	D	1	1	D	1	D	1	D	2	0	D	0	D	1	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 18	JAMPALA DIVYA	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	1	D	1	D	D	0	D	D	D	D	D	D	2	D	D
A - 19	JITESH SINDHARE	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 20	DEVENDRA THAKUR	1	D	1	D	1	1	D	1	D	1	D	2	0	D	0	D	0	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 21	KULDEEP	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 22	MAMTA THAKUR	1	D	1	D	1	1	D	1	D	1	D	2	0	D	0	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 23	MANISH KUMAR SINGH	1	D	1	D	1	1	D	1	D	1	D	0	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 24	MEGHA SONI	1	D	1	D	1	1	D	1	D	1	D	0	1	D	1	D	1	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 25	NEHA TRIPATHI	0	D	1	D	0	0	D	0	D	0	D	2	0	D	0	D	0	D	0	D	D	0	D	D	D	D	D	D	2	D	D
A - 26	NILESH S MUDLIAR	1	D	1	D	0	0	D	0	D	0	D	0	0	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 27	OMESH SUDHAKAR	1	D	1	D	1	1	D	1	D	1	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 28	PRACHI KATLAM	1	D	0	D	1	1	D	1	D	0	D	2	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 29	PRAVEEN ABHISHEK LAKRA	1	D	1	D	0	1	D	1	D	0	D	0	0	D	1	D	0	D	0	D	D	1	D	D	D	D	D	D	2	D	D
A - 30	RAJSHREE SEN	1	D	0	D	1	0	D	1	D	1	D	2	1	D	1	D	0	D	0	D	D	1	D	D	D	D	D	D	0	D	D
A - 31	RANJANA VERMA	1	D	1	D	1	1	D	1	D	1	D	0	1	D	1	D	0	D	1	D	D	1	D	D	D	D	D	D	2	D	D
A - 32	MANSI JAIN	1	D	0	D	0	0	D	0	D	0	D	2	1	D	1	D	0	D	1	D	D	0	D	D	D	D	D	D	0	D	D

Figure 1 Attendance of DELD (September)

Among all six working days, it is generally seen that attendance of students on Saturday is very poor. So there should be a system which tracks the attendance of students on every Saturday. In this experiment, name of the students are picked up who shall tend to be absent on first, second, third and fourth Saturday.

	DATE->	6	13	20	27	6	13	20	27	6	13	20	27	6	13	20	27	6	13	20	27	6	13	20	27	6	13	20	27
CRN	NAME OF CANDIDATE	1ST SAT/ BE	2ND SAT/ BE	3RD SAT/ BE	4TH SAT/ BE	1ST SAT/ CCWT	2ND SAT/ CCWT	3RD SAT/ CCWT	4TH SAT/ CCWT	1ST SAT/ CS	2ND SAT/ CS	3RD SAT/ CS	4TH SAT/ CS	1ST SAT/ DELD	2ND SAT/ DELD	3RD SAT/ DELD	4TH SAT/ DELD	1ST SAT/ MIII	2ND SAT/ MIII	3RD SAT/ MIII	4TH SAT/ MIII	1ST SAT/ PSLB	2ND SAT/ PSLB	3RD SAT/ PSLB	4TH SAT/ PSLB
A - 01	VARTIKA NAYAK	0	D	D	D	D	1	D	D	D	D	D	D	1	0	D	D	D	0	0	D	D	0	0	D
A - 02	AKASH AGGARWAL	0	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	0	D
A - 03	KUMARI PRIYA	2	D	D	D	D	1	D	D	D	D	D	D	1	0	D	D	D	1	0	D	D	0	1	D
A - 04	VIVEK KUMAR ROCKY	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 05	SHIKHA KUMARI JHA	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	0	D	D	1	1	D
A - 06	AKSHAY BELE	0	D	D	D	D	1	D	D	D	D	D	D	0	0	D	D	D	0	0	D	D	0	0	D
A - 07	ANJALI DHRUW	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	0	0	D	D	1	1	D
A - 08	ANURAG KUMAR SINGH	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	0	D
A - 09	G. NIVEDHA NADAR	2	D	D	D	D	1	D	D	D	D	D	D	0	1	D	D	D	1	1	D	D	1	1	D
A - 10	AVINASH MISHRA	1	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 11	CHINMAY JAIN	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	0	D	D	1	1	D
A - 12	DEEP SINGH	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 13	DEEPSHIKHA KURRE	2	D	D	D	D	1	D	D	D	D	D	D	1	0	D	D	D	1	1	D	D	0	1	D
A - 14	DEVIKA	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 15	DIKESH KUMAR	0	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	0	D	D	1	1	D
A - 16	DUSHAYANT KUMAR KOTHARI	2	D	D	D	D	0	D	D	D	D	D	D	1	0	D	D	D	0	1	D	D	0	1	D
A - 17	HANSHIKA AGRAWAL	2	D	D	D	D	1	D	D	D	D	D	D	1	0	D	D	D	1	1	D	D	0	1	D
A - 18	JAMPALA DIVYA	0	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 19	JITESH SINDHARE	2	D	D	D	D	0	D	D	D	D	D	D	1	1	D	D	D	0	0	D	D	1	1	D
A - 20	DEVENDRA THAKUR	2	D	D	D	D	0	D	D	D	D	D	D	1	0	D	D	D	1	1	D	D	0	1	D
A - 21	KULDEEP	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 22	MAMTA THAKUR	0	D	D	D	D	1	D	D	D	D	D	D	1	0	D	D	D	1	1	D	D	0	0	D
A - 23	MANISH KUMAR SINGH	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	0	0	D	D	1	1	D
A - 24	MEGHA SONI	2	D	D	D	D	1	D	D	D	D	D	D	1	1	D	D	D	1	1	D	D	1	1	D
A - 25	NEHA TRIPATHI	0	D	D	D	D	0	D	D	D	D	D	D	0	0	D	D	D	1	0	D	D	0	0	D
A - 26	NILESH S MUDLIAR	0	D	D	D	D	0	D	D	D	D	D	D	0	0	D	D	D	0	0	D	D	0	0	D

Figure 2 Attendance of students in each Saturday for all subjects in September

Figure 2 shows the attendance of student of all subjects in each Saturday. Here some column contains only D which means that classes was not held on that day. Following calculations are done:

$$SA = (S1+ S2+ S3+ S4+ S5+ S6) / 6$$

Where SA = Students attendance, S1 to S6 are subjects.
Value of SA will always be less than or equal to one. Name of the students can be picked up who have not attended on the first Saturday of each month, second Saturday of each month and so on.

3.2 Observation and Results

By doing the calculations in above subsection it has been observed that there are some students who are regularly absent on all the Saturday. Frequent student absentees in the forthcoming months can be easily predicted by looking the pattern of absenteeism in previous months. A threshold value of 0.75 is taken. Students having less than the threshold are displayed in figure 3. By looking the pattern of absenteeism in September month, it can be easily predicted the name of the students who may be absent in the same Saturday of the October or November month.

SEPTEMBER (< 0.75)

	FOR SAT 2
A - 01	VARTIKA NAYAK
A - 03	KUMARI PRIYA
A - 06	AKSHAY BELE
A - 13	DEEPSHIKHA KURRE
A - 16	DUSHAYANT KUMAR KOTHARI
A - 17	HANSHIKA AGRAWAL
A - 19	JITESH SINDHARE
A - 20	DEVENDRA THAKUR
A - 22	MAMTA THAKUR
A - 25	NEHA TRIPATHI
A - 26	NILESH S MUDLIAR
A - 29	PRAVEEN ABHISHEK LAKRA
A - 32	MANSI JAIN
A - 33	ROOPAK TRIPATHI
A - 38	VAIBHAV KUMAR BHARDWAJ
A - 39	YOGENDRA SAHU
A - 40	RAMAKANT BHAGAT
A - 42	FAHAD SAEED KHWAJA
A - 43	ESHITA JAIN
B-47	ABHIJEET
B-82	AAYUSHI A SHRIVASTAVA
B-85	AYUSH WADHAWA
B-87	RAVISH R. RAM
B-89	PRASANT RAWAL

OCTOBER (<0.75)

	FOR SAT 2
A - 01	VARTIKA NAYAK
A - 06	AKSHAY BELE
A - 07	ANJALI DHRUW
A - 14	DEVIKA
A - 21	KULDEEP
A - 25	NEHA TRIPATHI
A - 26	NILESH S MUDLIAR
A - 29	PRAVEEN ABHISHEK LAKRA
A - 30	RAJSHREE SEN
A - 32	MANSI JAIN
A - 33	ROOPAK TRIPATHI
A - 40	RAMAKANT BHAGAT
A - 41	DEVIKA KULKARNI
A - 42	FAHAD SAEED KHWAJA
A - 44	ANSHI JAIN
B-47	ABHIJEET
B-56	DERARUN DAS
B-64	KHEMRAJ THAKUR
B-67	MANISH KUMAR KURREY
B-75	AVADESH ANAND AGARWAL
B-85	AYUSH WADHAWA
B-87	RAVISH R. RAM
B-93	ALISHA RASMI KHALKHO

NOVEMBER (<0.75)

	FOR SAT 2
A - 01	VARTIKA NAYAK
A - 03	KUMARI PRIYA
A - 06	AKSHAY BELE
A - 07	ANJALI DHRUW
A - 13	DEEPSHIKHA KURRE
A - 15	DIKESH KUMAR
A - 18	JAMPALA DIVYA
A - 20	DEVENDRA THAKUR
A - 26	NILESH S MUDLIAR
A - 29	PRAVEEN ABHISHEK LAKRA
A - 32	MANSI JAIN
A - 33	ROOPAK TRIPATHI
A - 35	SHUBHAM RAWANI
A - 38	VAIBHAV KUMAR BHARDWAJ
A - 40	RAMAKANT BHAGAT
A - 42	FAHAD SAEED KHWAJA
A - 44	ANSHI JAIN
A - 95	PURVA MOON
B-47	ABHIJEET
B-75	AVADESH ANAND AGARWAL
B-80	THAKUR PRASAD VISHAL
B-83	ASHWARYA S. PILLAI
B-85	AYUSH WADHAWA
B-86	MOHIT AGARWAL

Figure 3 students who are in alert zone

4. CONCLUSION

By the above experiment it can be concluded that by tracking the attendance pattern, the performance students can be easily improved. Real time data can be used in academic advising early warning system. The persons who are associated to students can identify at - risk students.

REFERENCES

[1] Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access*, *2*, 652-687.
[2] Rajeswari, S., & Lawrance, R. (2016, January). Classification model to predict the learners' academic performance using big data. In *Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on* (pp. 1-6). IEEE.
[3] West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance Studies at Brookings*, *4*, 1-0.
[4] Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016, April). Big data application in education: dropout prediction in edx MOOCs. In *Multimedia Big Data (BigMM), 2016 IEEE second international conference on* (pp. 440-443). IEEE.
[5] Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting student performance using personalized analytics. *Computer*, *49*(4), 61-69.
[6] Rabelo, T., Lama, M., Amorim, R. R., & Vidal, J. C. (2015, October). SmartLAK: A big data architecture for supporting learning analytics services. In *Frontiers in Education Conference (FIE), 2015 IEEE* (pp. 1-5). IEEE.
[7] Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9-20.

[8] Bhat, A. Z., & Ahmed, I. (2016, March). Big Data for institutional planning, decision support and academic excellence. In *Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on* (pp. 1-5). IEEE.
[9] Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.
[10] Wong, G. K., & Li, S. Y. (2016, June). Academic performance prediction using chance discovery from online discussion forums. In *Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual* (Vol. 1, pp. 706-711). IEEE.