

Data Mining- Detect Exam Cheating Pattern

Mason Chen

Milpitas Christian School

San Jose, CA, USA

mason.chen.training@gmail.com

Timothy Liu

Lynbrook High School

San Jose, CA, USA

timothys.new.email@gmail.com

Abstract

This project will demonstrate on how to apply Data Mining on detecting the exam cheating pattern. JMP software was utilized to analyze the multi-correlation among 75 students (sit on 25 different tables) taking an assessment exam with multiple choices. Three students from the same table will take the same exam but with different sequence of questions. However, students were smart to synchronize the question sequence immediately. In order to detect the cheating pattern, multivariate statistics was used to determine whether there was any association among the students from the same table. Hierarchical Clustering and Dendrogram Tree can identify the grouping affinity behavior related to exam cheating pattern. Team also used Heat Map to visualize the cheating pattern. To improve the prediction confidence, team also selects the top 20% difficult questions to increase the detection signal-noise ratio to minimize both Alpha and Beta Error below 1%. At such lower level, students won't defend their cases statistically. Based on the results, three tables were identified with cheating patterns. The predictive model was very powerful to analyze complicated cheating patterns.

Keywords

Big Data, Data Mining, Clustering, Heat Map, Predictive Model

Acknowledgements

Dr. Charles Chen

Biography

Mason Chen is currently a student in the Milpitas Christian Middle School. Mason has certified IASSC (International Associate of Six Sigma Certificate) Lean Six Sigma Yellow Belt, Green Belt, and Black Belt Certificates. He has also certified IBM SPSS Statistics Certificate. He also won the 1st Place Award on the Mental Math and Abacus Math contests in the North California Region. Mason Chen is familiar with Six Sigma DMAIC, DMADOV, Lean Production, Minitab, SPSS Statistics, SPSS Modeler CRISP Data Mining, AP Statistics, and JAVA tools. Mason got invited to present his five ASA team statistics projects for 90mins in the local ASQ Statistics and Reliability Group.

Timothy Liu is currently a student in the Lynbrook High School. He has been learning JAVA Programming and AP Statistics. He has certified an IBM SPSS Statistics Certificate and also familiar with IBM SPSS Modeler Data Mining Software. Timothy has also received a solid Six Sigma Yellow Belt Training this summer and familiar with

DMAIC Six Sigma Problem Solving Methodology. Timothy got invited and presented his JAVA Project in the Local American Society for Quality (ASQ) Statistics and Reliability Group.