EMOTION ANALYTICS WITH PROCESS MINING

Prajin Palangsantikul[1], Parham Porouhan[2], Nucharee Premchaiswadi[3], Wichian Premchaiswadi[4]

[1,2,4]Graduate School of Information Technology, Siam University, Bangkok 10160, Thailand

[3]College of Creative Design and Entertainment Technology, Dhurakij Pundit University, Bangkok 10210, Thailand

Email: prajin.pal@siam.edu[1], parham@siam.edu[2], nucharee@dpu.ac.th[3], wichian@siam.edu[4]

DOI: https://doi.org/10.22452/mjcs.sp2019no2.7

ABSTRACT

This research builds on the intersection of 'process mining' and 'emotion analytics' in order to discover and investigate the emotional patterns of students during the StudentLife project based on data collected by smartphones through PAM application (i.e., Photographic Affection Meter). The main objective of the study is to analyze and predict the relationships (or continuity) amongst 16 common emotional indicators based on the 'Russel's Circumplex Affect Grid' and by means of Fuzzy Miner (supported by Disco Fluxicon) and Dotted Chart Analysis (supported by ProM) process mining tools and techniques. Accordingly, the current work is divided into two main parts. In the first part, a pre-processing (or data preparation/cleansing) approach via Python programming was done in order to change the format of the initially collected event logs from JSON to the appropriate format/structure. In the second part, the emotional datasets were analyzed using the above-mentioned techniques. To do this, new groups/clusters of contexts were categorized and pre-defined. The third part of the study deals with data interpretation and discussion of the obtained findings. The proposed/applied approach was capable of providing "frequency-based" models/graphs of the students' behavior, both before and after the experiment, in terms of 5 categories: "Minimal, Minor, Moderate, Moderately Severe, and Severe" depression-related emotional trends ranging from Low Severity (NA) to High Severity (PA). This research provides groundwork for further and future studies.

Keywords: Process Mining, Smartphone, Fuzzy Miner, Dotted Chart Analysis, Emotion Intelligence, StudentLife project, ProM, Disco Fluxicon

1.0 INTRODUCTION

The current society is highly competitive in all aspects, especially at study and work. One in five adults in the United States have mental health problems which means more than 40 million people and the rate of adolescents with severe depression has increased from 5.9% in 2012 to 8.2% in 2015. Moreover, 76% of youngsters with severe depression were left without treatment [1]. Dealing with teenagers, this is the age that has changed dramatically, both physically, mentally, and socially, so teenagers normally undergo and face a variety of stress in this period of time. In addition, we are rapidly witnessing a social change in today's society values, and culture, all influencing the lives of adolescents who have high expectations, especially dealing with crucial issues like education and studies. These factors are the main causes of stress amongst 'today's teenagers' in their daily lives compared with the teenagers in the past. Up to this point, many studies have focused on the field of 'emotional education' in order to find out how to check the mood of the person at risk, or to find signs of emotion-induced diseases, in order to address or prevent the diseases with proper interventions on the right time, because emotion is a major indicator of a person's mental health. By knowing the overall mood (or the current mood) of a person earlier, this can significantly play an important role in prevention of some mental-related health problems. In recent years, the popularity of mobile apps have corresponeded with rapid growth in the use of smartphones. Therefore, there is an increasing opportunity for researchers to access and collect more sophisticated types of mental health personal data (event logs) than ever.

The main emphasis of the study is to discover and evaluate the state of a person's emotions using process mining techniques with the purpose of analyzing the emotions of students from data previously collected by a smartphone. To do this, the research is divided into five main sections.

- Section 2 emphasizes on the related works and research previously conducted in the areas/fields of "Process Mining and Emotion Analytics" in addition to "Emotion Analytics".
- Section 3 focuses on the research methodology while providing more information about the preparation of the emotional data, which was collected and stored from the PAM (Photographic Affection Meter) application [2] installed on the smartphone during the StudentLife project (i.e., so-called pre-processing or cleansing of data phase/stage), as well as the provision of more details about the two adopted process mining techniques.
- Section 4 includes the analysis of data (and findings of the study) through several Process Mining techniques [3] as an effective method to discover emotional patterns of the students. Such techniques are useful in finding in-depth knowledge based on datasets or event logs collected by an Information System. Process Mining tools can facilitate and streamline the visualization of ongoing activities and processes based on the actual data. Most of the Process Mining techniques are used in business assessments that have a well-defined operational and data storage model. In this research, the main idea was to apply process mining techniques in order to discover and evaluate the emotional patterns of the sample students.
- Section 5 includes the discussion and explanation of the findings based on the results obtained from the ProM and Disco Fluxicon [4] process mining tools; yielding a model of emotional overview for the students' emotional behavior throughout the assigned project.

2.0 RELATED WORK

Over the past two decades, great advances in the field of Information Technology have evolved into the emergence of a rapidly growing digital age. Every second there is a huge amount of information being generated or passed through an Information System. The discovery of knowledge in traditional databases has been developed into a new technology called Process Mining, which is a technique for knowledge discovery or conformance checking. Process Mining is based on a model-based simulation of the actual datasets (in terms of event logs) collected from various Information Systems. Currently, different forms of data with different structures and formats can be easily found and accessed from versatile venues and systems. Sometimes the large amounts of raw data are not used effectively (or efficiently) for analytical purposes, but process mining techniques can fill this gap. On the other hand, this raw data has the potential to be extracted for in-depth analysis of latent knowledge in terms of datasets so-called event logs. As an instance, such data can be used to verify (or assess) the consistency of the processes with a master model, to detect and discover bottlenecks, or to predict operational problems. Some systems, such as; monitor workflow management systems or transaction logs; enterprise resource planning systems; can also be used and applied in process mining. Other possibilities range from gaining insights and generation of mdoels can be used in order to check any deviation or violation of the rules within a system to get insights to discover new process models, or to check and compare the consistency between the ongoing tasks and a model. Such capabilities can have profound benefits, especially when managers, researchers or scholars look to improve the efficiency of a system using process mining techniques in various tasks or scenarios/situations [3].

2.1 Process Mining and Emotion Analytics

Traditional data mining techniques focus on data referencing or simple modeling even though there is no complete picture of the study in the model ready to be analyzed. To be able to better (and further) analyze datasets of this type, a new process management field so-called process mining emerged. Process mining focuses on developing toolkits and intelligent techniques with the purpose of extracting knowledge from event logs that are collected by an information system. As mentioned earlier, one of the process mining dimensions is to discover knowledge from datasets. To do this, there are many algorithms that can automatically generate a process model. The process model representation usually comes in the form of a mathematical model. It supports the simultaneous operation of sequential and alternative behaviors (e.x., Petri nets, Heuristic or Fuzzy Miner). In other words, process mining is the missing link between business process management (BPM) and non-process analysis techniques, such as data mining, Machine learning, and business intelligence.

In this paper, two process mining tools, Disco Fluxicon [4] and ProM [5], were applied in order to find knowledge from the student's emotional dataset. Disco Fluxicon is a process mining tool that is fully functional and is a trademark of Fluxicon Company. ProM is another popular process mining open source software that contains and supports more than 600 plug-ins. In the current work, the data was analyzed and an evaluation of the experiment through the emotional model was provided. Unfortunately, the majority of research done on the topic of process mining has been focused on business scenarios, due to the fact that in a business setting it is easier to collect event logs containing activities with a systematic and structured process. However, very few studies have focused on the implementation of process mining tools/techniques to analyze human behavior. Especially, the emotional aspect of students in an academic setting. One of the reasons is the difficulty to store, collect and gather datasets with appropriate format and structure from such settings. Therefore, this research applied some of the process mining techniques, previously used for business management purposes, in an academic environment. For example, Bozkaya and Gabriels and Werf [6] used a methodology to perform process diagnostics based on process mining. Given an event log of an information system within a Dutch governmental organization, the applied process diagnostics was capable of giving a broader overview of the organization's process within a shorter period of time. Accordingly, by applying the process diagnostics methodology, several perspectives of the process activities were highlighted. The outcomes included the control flow perspective, the performance perspective and the organizational perspective. Bozkaya and Gabriels and Werf [6] applied the above-mentioned methodology on a case study for a Dutch governmental organization.

Until today, the only similar work in which the same ProjectLife data was used, is a study done by SmartGPA Research [7]. However, in SmartGPA Research [7], data analytical methods and techniques were used in order to predict students' emotional behavior trends; while no process mining approaches/techniques have ever been applied on the StudentLife datasets/event logs [8]. In their work, they focused on the relationship between the final GPA of the students with numerous variables (e.x., such as type of personality, student's mental health, stress affect, length/duration of conversation, students' active hours, duration of partying, duration of studying, etc.) which were then statistically analyzed and investigated. Although SmartGPA Research [7] did not apply any process mining techniques, their work was the main motivation of this research, since the same datasets/event logs were used and studied in this research. One of the advantages of the current work compared with the work done in [7] is that, the process mining techniques can significantly streamline the human behavior analysis process, making it easier to observe and investigate the main concerns of mental health, mental (or stress-related) problems, social relationships, and their relations to overall performance of students. Accordingly, in this research, student's emotional patterns were identified and studied based on a series of datasets previously collected from Smartphones (and freely shared and accessible to public) within the StudentLife project. Fortunately, the number of students who are using Smartphones to communicate with each other is day by day increasing, and this was the first time that process mining tools and techniques were applied on StudentLife project event logs [8].

In other work done by Zig Websoftware [9], the process mining techniques were used in order to better visualize, illustrate and gain some insightful information regarding the feelings of the process managers of housing associations based on data containing 4000 days of vacant properties collected within a 6-month period of time. The Zig Websoftware [9] experiment was conducted by Disco Fluxicon process mining software in an automation platform setting/venue. The results of their work could significantly help the housing managers to make faster and further amendments in the process. Moreover, by using Disco Fuzzy Miner process mining techniques and by automation and simulation of specific tasks within the apartment allocation process of WoonFriesland, the managers could better benchmark and optimize the overall investment rate and cost savings, leading to considerable reduction of the vacancy rate.

2.2 Emotion Analytics

In work done by [10], the impact of various arousal-inducing manipulations on dual task performance was implemented and investigated. The study was based on the Easterbook Cue-Utilisation hypothesis [10] which aimed to investigate whether an increase in the extent of experienced arousal leads to decreased understandability /comprehension of a topic. In their work, the extent of the experiencing arousal had a high impact on both behavioral and intellectual learning. Their results showed that emotional arousal may act consistently so as to decrease the range of cues an organism uses, and such reductions can organize or disorganize an action in accordance with the behavior observed.

In research done by "J. A. Russell" entitled "Affect Grid: A single-item scale of pleasure and arousal" [11], the Affect Grid was designed and introduced as a quick means of assessing human affections/emotions in terms of pleasure–displeasure and arousal–sleepiness dimensions. The proposed "Affect Grid" is potentially suitable for any study that requires judgments about human affections/emotions, either in forms of descriptive or subjective approaches. The scales and dimensions used in PAM application were previously proved to have adequate reliability (i.e., in terms of both convergent validity and discriminant validity) based on the several studies in which college students used the "Affect Grid" to describe/express their current status of their emotions [11]. The study was based on the Russell's Affect Grid, which is sometimes called Circumplex's Affect Grid.

In this paper the PAM data was used to collect and measure the psychosocial emotions of students based on the PANAS Scales ratings. This approach originated in the work done by Watson and Clark and Tellegen [12]. In their work, both Positive and Negative affect dimensions were defined and used as two important (and independent) variables. Up until this point, this work has been the most comprehensive research that considers both Positive and Negative emotions in terms of PA (Positive Affection) and NA (Negative Affection) values in order to statistically assess and measure their relationships with some social activities which were significantly correlated with 'perceived stress' and 'circadian rhythms' such as: 1) Sleep 2) Hormone 3) Metabolism 4) Body temperature 5) Other factors/variables. To do this, a set of mood-related scales were generated with the intention of measuring the extent of the above-mentioned variables. In order to provide rather straightforward and concise "Positive Affect" and "Negative Affect" scales (with high degree of validity, reliability and convenience), Watson and Clark and Tellegen [12] applied two 10-item mood-related scales that made up what they called PANAS; a Positive and Negative Affect Schedule. Although, the way they defined and investigated the mood scales was interesting, their work merely focused on the statistical analysis of the variables, while investigating correlations and dependencies between the variables. Nevertheless, no data mining or process mining techniques were used in their research. The findings of their work showed appropriate levels of consistency, stableness and correlation between the mood scales during a 2-month experiment study. Subsequently, normative values as well as the external/factorial proofs of convergent and discriminant validity coefficients in support of the mood-related scales, were discussed and presented. The reason why the work became the main motivation of this study is because of the fact that Watson and Clark and Tellegen [12] offered a two dimensional grid[1] which included both positive and negative emotional values in a reliable, accurate, and effective approach. In other words, quite interestingly, the effects of both positive and negative emotions (and their relationships) were measured on the overall performance of the students.

Another study was based on the Yerkes-Dodson Law [13] which aimed to investigate the relationship amongst two variables of "stimulus strength" and "habit-formation". The research aimed to study the change of the Yerkes-Dodson law during the years 1908 to 1994. In their work, they investigated the relationship between the concepts "Anxiety", "Level of Stress" and "Efficiency of Performance" through Hebb/Yerkes-Dodson hybrid. Their proposed model indicated the fact that the extent of performance increases with physiological or mental arousal, but only up to a limited/specific point. According to the Yerkes-Dodson Law, when the levels of arousal become too high (or overly aroused), then the efficiency of performance decreases, usually followed by emotional disturbance, disorganization, feeling frenzy or panic, etc. Ultimately, in their work, the relationships amongst "Arousal Level" and "Performance" were illustrated graphically as a bell-shaped curve which increases and then decreases with higher levels of arousal.

Another study was based on the Arousal Theory [14], which stated that humans are motivated to perform certain behaviors. To maintain/reach an optimum level of arousal, when the level of alertness is lower, the level of stimulated arousal should be increased. In the same way, when the level of alertness is too high, the level of stimulated arousal should be pulled down. Each person has a different level of alertness. However, a good example of this relationship occurs when we wake up from sleeping. Researchers have used many methods to measure the extent of relationships amongst Arousal and Alertness of the human body (i.e., normally by measuring brain waves, heartbeat muscle contraction, or through state of the organs, etc.). While sleeping, the level of alertness in a human body will be minimal. However, if anything shocking or accidental happens, then the extent of Arousal in the human body increases.

Another study [15] was based on a two-dimensional emotional Positive Activation (PA) and Negative Activation (NA) structure chart (i.e., including human dimension the stimulus-driven dimension) proposed by David Watson et al. In their work, the main idea was based on the fact that PA and NA can be very useful explanatory variables that have the potential to provide better clarifications in regard to some important affect-related constructs such as sleeping habits/hours, mood disorders and circadian rhythms; and as building blocks of a broader phenomena which is called the human's bio-behavioral system. One of the advantages of their work was that David Watson et al.'s emotional chart aimed to group the categories of emotions in an extremely simple (and easy to understand) emotional model.

Robert Plutchik [16] proposed a three-dimensional circumplex model which aimed to investigate the relationship between emotional constructs based on a colorful wheel so-called "The Nature of Emotions or Emotional Wheel". In their work, the vertical dimension was an indicator of the extent of "intensity", while the circle was an indicator of the degrees of similarity between the emotional concepts. Accordingly, eight sectors were defined and designed in order to represent eight primary emotional dimensions illustrated and divided into four pairs of opposites. Robert

[1] For more information about the work done by Watson and Clark and Tellegen [9] please check out the link below: https://pdfs.semanticscholar.org/7124/10f2b20c831678f71db35ec01ba1b38bc842.pdf

Plutchik [16]'s proposed model could predict the occurrence of eight types of emotions when facing a shocking or extremely exciting experience as the following: "fear, surprise, sadness, hatred, anger, anticipation, joy and acceptance". However, the occurrence of each of these eight emotions may vary according to the level of experienced 'emotional intensity'.

In another study entitled "Emotion Mining Research on Micro-blog" [17], the effect of Human Emotions on 3 elements of "physical health", "mental health", and "human performance" were studied and investigated with high degree of accuracy and through a user-friendly approach while defining new emotional paradigms and indicators. In their work, in order to discover and specify the emotional trends based on a large set of textual datasets, and in order to predict the feelings associated with the micro-blog data, in this paper, the following actions were done and implemented: (i) The emotional words/terms were identified and categorized. (ii) An emotional words weight dictionary (WD) including 1342 words/terms was developed. (iii) An additional negative words dictionary (NWD), containing the words carrying negative emotions, was pre-defined and constructed. (iv) Another interjection words dictionary (IWD) was defined and constructed. (v) Both of the words dictionary (DWD) and interjection words dictionary (IWD) were allocated specific "weights" and coefficients". Accordingly, a statistical classification was performed on 2213 micro-blog items, with the intention of discovering/identifying those words/terms in which contain and express positive and negative emotions according to their allocated weights. Subsequently, a calculation of the weights, for each word/term of the micro-blog data took placed and executed with an accuracy rate of 80.6%.

Research done by Stress State (EEG) [18] focused on those types of emotions that are associated with hormones by proposing a novel "stress recognition system" which works based on multi-modal bio-signals. In this study [18] a new labelling process of electroencephalogram (EEG) signals for recognition of the emotional stress state (EEG) was used and applied. To achieve this, and in order to select the appropriate type of EEG channels, the cognitive model of brain activity in accordance with emotional stress was used and investigated in detail. Accordingly, after pre-processing the electroencephalogram (EEG) signals, both Linear and Non-Linear methods were used with the purpose of identifying and extracting the most important stress-related emotional EEG parameters with an accuracy rate of 82.7%. Although, this work was technically interesting and the relationship between brain hormones and stress-related emotional activities was studied and discussed, their work merely contained statistical calculation of the relationships amongst the received signals and did not focus or apply any data mining or process mining approaches/techniques.

As discussed earlier, the main objective of the research was to find patterns of students' emotions based on StudentLife datasets previously collected via PAM mobile application [2], which is a new tool to measure human emotions while carrying/using their cellphones. Using PAM, users can choose from a variety of photos, which appear in the application and in such a way that the selected photo best suits/explains their (current) emotion. Previous studies have shown that the PAM, which takes a few seconds to process, is suitable to demonstrate a great deal of authenticity (and reliability) dealing with the users' emotional exploration, and therefore, it is convenient for frequent sampling purposes. However, prior to getting/collecting reliable data, it is important to understand some basic psychological knowledge about how the PAM application works psychologically in order to identify and interpret the emotional state of a user.

Many emotional research studies have focused on measuring the extent of stress such as work done by [19] entitled "Mobile Stress Tester: Using Images", while the work is capable of quickly measuring the level of stress in real time with high coefficient of reliability and accuracy. The work presented in [19] runs based on a user-friendly application which provides an enjoyable experience for users who have tried it. Most importantly, the developed application works effectively for measuring the level of stress scientifically and compatible with a research approach.

Moreover in work done by [19] entitled "Positive and Negative Affect and Arousal: Cross-Sectional and Longitudinal Associations with Adolescent Cortisol Diurnal Rhythms", the relationship between the extent of Arousal and (Positive or Negative) Emotions was studied and elaborated. The study [20] showed that 'positive emotions' accompanied with 'high arousal' leads to the reduction of 'Cortisol Hormone' in human body. This is what is called a "Cortisol Emotional Experiences" in medicine and it normally happens when experiencing extreme emotional pressure such as: anger, sadness, loneliness, or depression. The phenomenon is normally associated with an increase of hormone cortisol in human body's blood.

As a result, in this study a set of metadata analysis based on preliminary studies was followed and implemented in such a way that the mood of a user was certainly found to have had an impact on various mechanisms that influenced the overall performance dimension.

3.0 METHODOLOGY

The main objective of the research was to study the patterns of emotions amongst the students during the StudentLife project at Dartmouth College, where the data was collected from the PAM application installed on the students' smartphones. The data consisted of 40 samples. Within the data preparation phase, the Python programming language was used so as to prepare the data for further analysis and investigation. This step is described in detail in Section 3.1. In this section, the "Circumplex Affect Grid" was used as an initial model to specify 16 main emotional parameters/indicators. In order to simplify the conceptual framework of the students' emotions, the total number of parameters was reduced to 4 indicators.

3.1 Data Preparation

StudentLife Dataset consists of two main types of data: Sensor Data (Automatic) or Passive Data and EMA Data (Active). PAM data is part of the EMA data. It stores the information obtained from the image selection on the PAM mobile app. PAM, is a 1-16 score that maps to PANAS PA and reports as Valence, Arousal [2]. The data is then sent to the server to collect each student in a JSON file format. The numbers match the selected image and the time at that moment. Accordingly, the PAM data will be ready for use in (process) mining analysis and discovery models. As mentioned earlier, in this research, the Python language was initially used to convert the PAM data into the appropriate format with desirable structure and needed data.

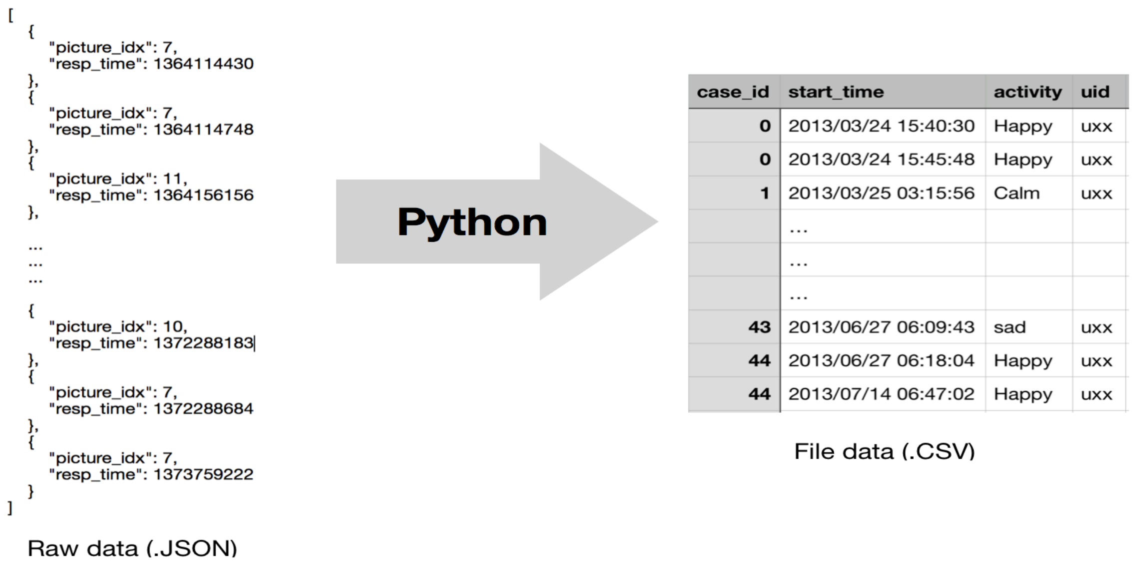


Fig. 1: A screenshot of the data conversion/preparation process through Python language

The PAM data set consisted of 40 files (or cases). Then the numeric values in the field were mapped. As shown in Fig. 1, the columns named "Picture_idx" in JSON format, which included the emotion list containing the type of emotions as well as their numeric values, were converted into CSV format while indicating the exact "Emotion Type" of the performed actions. In the same way, the columns named "resp_time" in JSON format, which included the Unix time information, were converted into CSV format while indicating the Standard time format. After the data preparation, data conversion, data cleansing and data filtering steps, a combination of both "Frequency-Based (Performance Analysis)" and "Directional Change (Trend Analysis)" graphical simulations/ illustrations of the "depression-related emotions" in accordance with "Affect Circumplex Model", for both pre-test and post-test scenarios, using "Fuzzy Miner (via Disco Fluxicon)" and "Dotted Chart Analysis (via ProM)" process mining techniques and tools, were implemented, analyzed and represented. Fig. 2 provides more details and information about the research methodology and two adopted mining techniques.

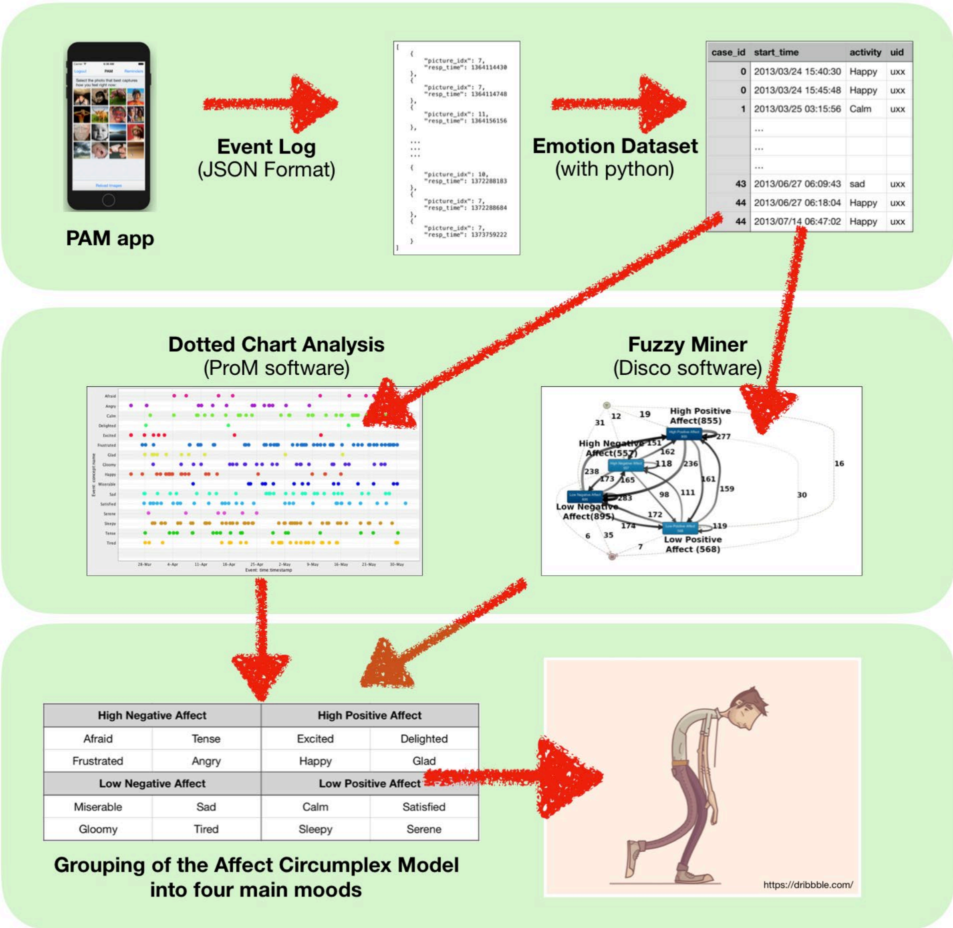


Fig. 2: An illustration of the research methodology providing details and information about "data preparation phase" as well as the two adopted mining techniques

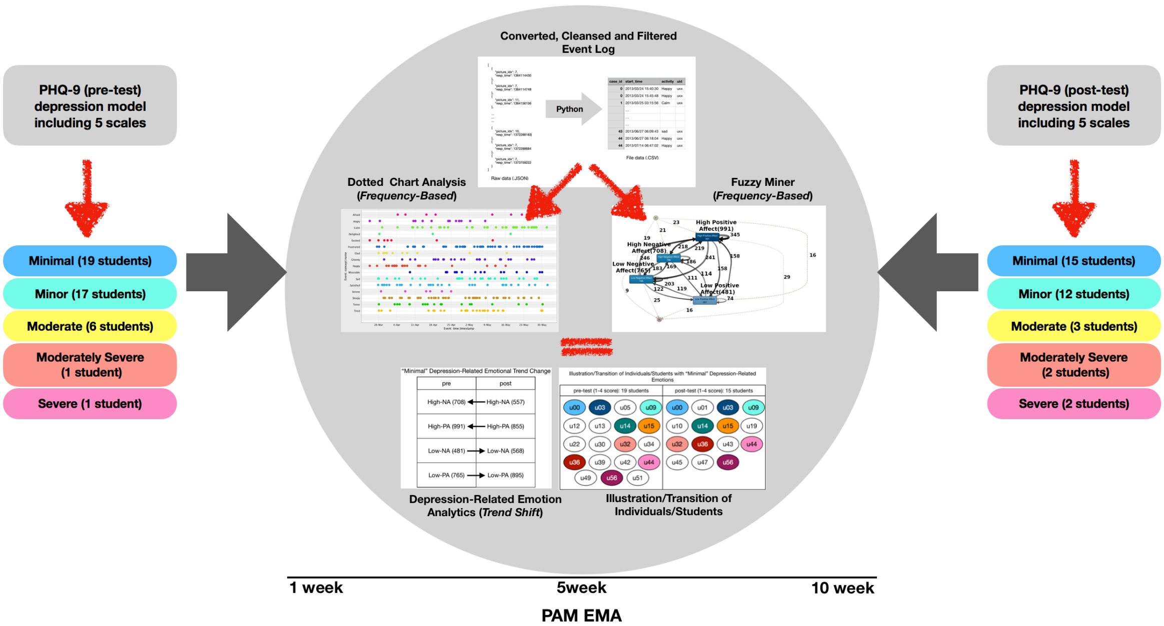


Fig. 3: A holistic view of the research methodology providing more details and information about the categorization/grouping of the depression-related emotions in addition to their relationship with the applied mining techniques

3.2 Fuzzy Miner (via Disco Fluxicon process mining platform)

Fuzzy Miner [21] is one of the new discovery algorithms developed by Fluxicon. In 2007, Christian W. Günther, the co-founder, was the first person who developed the Disco Fuzzy Miner algorithm[2] in order to solve problems associated with those types of activities in which containing complex or unstructured process instances. The rationale idea behind the the Disco Fuzzy Miner algorithm was to simplify the format in an interactive way. Consequently, as in this research, we were dealing with a very complex and unstructured type of emotional data. The Disco Fuzzy Miner algorithm was used to simplify the visualization of the StudentLife data. To provide more details about how the results (outcomes) of the Fuzzy Miner technique are interpreted, it is important to mention that the "Rectangular Boxes" or "Nodes" in the generating graphs typically represent the "Process Instances" (i.e., such as 16 emotional indicators or 4 main depression-related clustyers/scales). The illustrated "Numbers" inside the "Rectangular Boxes" normally stand for the "Frequency" (or number of times) those "Process Instances" have been executed within the collected event log(s). Accordingly, the "Arrows" from one node to another mode represent the relationship amongst those nodes, or the number of times Node 1 is followed by Node 2, subsequently. In general, the Fuzzy Miner approach provides useful insights and information regarding: (1) The sequence or order of the performed "Process Instances". (2) The most important "Process Instances" in terms of the frequency or number of times they have been executed. (3) The relationship and dependencies between the performed "Process Instances" according to the collected event log(s).

3.3 Dotted Chart Analysis (via ProM 6.2.4 process mining platform)

The Dotted Chart Analysis algorithm is part of the official distribution of the ProM process mining toolkit. ProM Dotted Chart Analysis algorithm is ideal for dealing with less structured processes that exhibit a large number of unstructured and conflicting behaviors. Consequently, given that this research used a very complex and unstructured spaghetti-like type (i.e., raw data) of emotional StudentLife datasets, to provide more details about the results (outcomes) it is important to mention that Dotted Chart Analysis[3] is based on the generation of charts which are very similar to Gannt charts.

The Dotted Chart Analysis technique is capable of representing a spread of events (or process instances) of an event log (or dataset) over a specific period/range of time. In other words, the main idea behind the technique is to illustrate dots along with the time supported by popular ProM process mining tool/platform. Accordingly, a dot plotted by ProM can illustrate an explicit event rooted in its relevant event log. The charts generated by ProM Dotted Chart Analysis technique contain 2 orthogonal dimensions: (1) "time dimension" along the horizontal axis, and (2) "event type" dimension (i.e., such as tasks, activities, originators, component types, instances, or any other additional attributes) along the vertical axis. Accordingly, in this study, both Disco Fuzzy Miner and ProM Dotted Chart Analysis algorithms were used in order to simplify the visualization process of the data in a more effective and efficient way.

---

[2] For further details about the main principles, mathematical foundations and algorithm of the Disco Fluxicon Fuzzy Miner technique, please refer to the following sources:
https://fluxicon.com/disco/files/Disco-User-Guide.pdf
https://fluxicon.com/disco/files/Disco-Tour.pdf
Aalst WMP (2011) Process Mining: discovery, conformance and enhancement of business processes. Springer, New York

[3] For further details about the main principles, mathematical foundations and algorithm of the ProM Dotted Chart Analysis technique, please check out the below sources:
http://www.processmining.org/online/dottedchartanalysis
M. Song and W. van der Aalst (2007), "Supporting process mining by showing events at a glance," in WITS 2007
Aalst WMP (2011) Process Mining: discovery, conformance and enhancement of business processes. Springer, New York

4.0 RESULTS AND FINDINGS

4.1 Initial "Emotion Model" based on 16 Emotional Indicators

Emotions are complex and watching the continuity of emotions on a daily basis is difficult to follow; as each day emotions change according to the environment and the impact of the things that affect them, resulting in mood changes. Fig. 4 shows the values of '16 emotional indicators' collected from the PAM dataset over a 64-day period for 43 students/members in the form of Absolute Frequency and Relative frequency. By looking at the results in Fig. 4, it is obvious that the emotions "Happy", "Tired", "Sleepy" and "Angry" allocated the highest Relative Frequency of 16.33%, 8.75%, 8.67% and 8.36% throughout the StudentLife data.

Activity	▲ Frequency	Relative frequency
Happy	1,369	16.33 %
Tired	733	8.75 %
Sleepy	727	8.67 %
Angry	701	8.36 %
Frustrated	685	8.17 %
Satisfied	682	8.14 %
Calm	672	8.02 %
Sad	472	5.63 %
Excited	440	5.25 %
Glad	356	4.25 %
Tense	310	3.7 %
Delighted	302	3.6 %
Gloomy	274	3.27 %
Serene	230	2.74 %
Afraid	214	2.55 %
Miserable	214	2.55 %

Fig. 4: A representation of the values related to 16 emotional indicators based on Frequency and Relative Frequency from PAM StudentLife event log

Although Fig. 4 is capable of indicating the maximum/minimum number of emotional values received by the PAM, what is lacking is the extent of relationship (and dependency) amongst the emotions as well as their correlations. PAM raw data was not capable of providing details in regard to the emotional relationships or continuity of the emotions. Similarly, PAM raw data could not provide more details about how emotional processes are followed and executed during the StudentLife project. Nevertheless, process mining tools and techniques can be useful to provide answers for the above-mentioned questions and ambiguities.

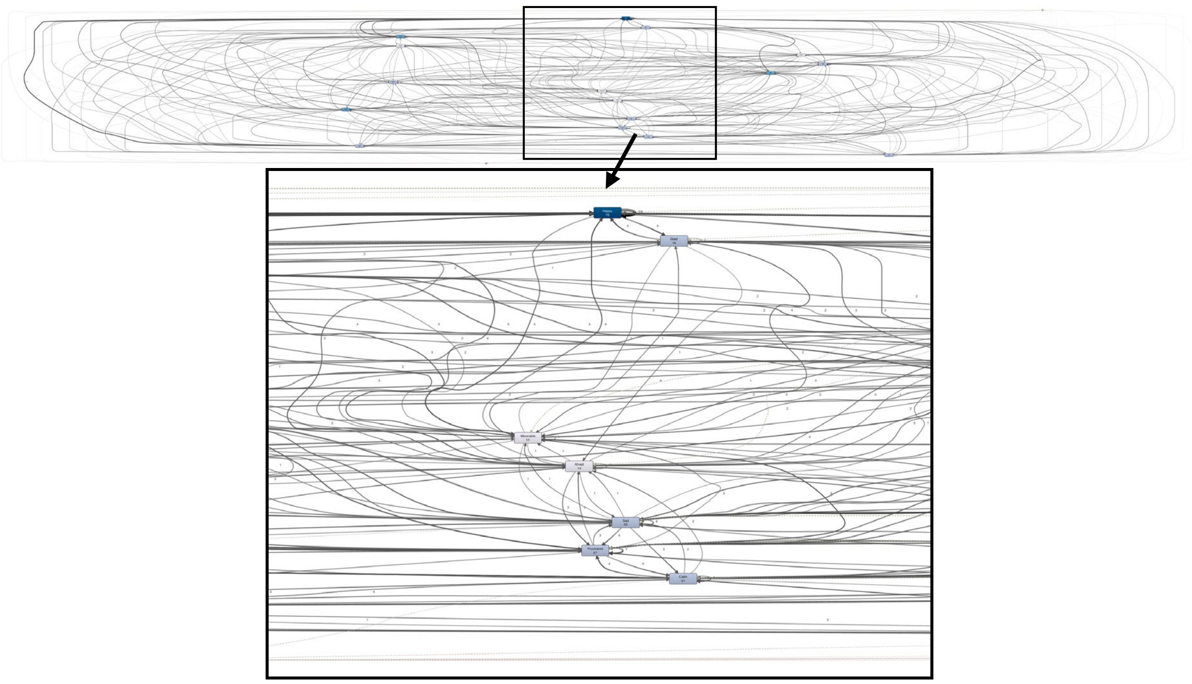


Fig. 5: A screenshot of the generated “spaghetti-like” Disco Fuzzy Miner process model, for 100% of Activities and 100% of Paths, without any threshold simplification cutoff, based on the PAM emotional data during StudentLife project

Fig. 5 shows a holistic (overall) process overview of the raw StudentLife event log data based on the PAM. Moreover, Fig. 5 illustrates the relationship amongst all of the emotional data indicators (values) by using the Disco Fluxicon process mining technique while displaying 100% of Activities and 100% of Paths, without any threshold simplification cutoff based on the PAM emotional data during StudentLife project. Although, Fig. 5 is capable of indicating the extent of relationships amongst the emotional indicators occurring within a 64-day project experiment, the "spaghetti-like" nature of the resulting process model, which was associated with less structured processes, made it difficult to comprehend, interpret and analyze the resulting dependencies between the tasks. The problem was that, without any threshold cutoff simplification of the processes, the generating Disco Fuzzy Miner process map displays too much detail in such a way that makes it difficult to interpret the data, or gain some insights about the PAM emotional event logs. To deal with this issue, a threshold simplification cutoff was implemented on the Disco Fuzzy Miner pre-default data settings. Using this approach makes it possible to reduce and simplify the number of activities or the relationship amongst the activities (i.e., paths). Fig. 6 shows the resulting Disco Fuzzy Miner model for all of the emotional activities (100% Activities) with 50% of Paths, respectively. Moreover, by following a Path Reduction method, it is more convenient to investigate the emotional relationships amongst the indicators (values) in a more specific and understandable manner.

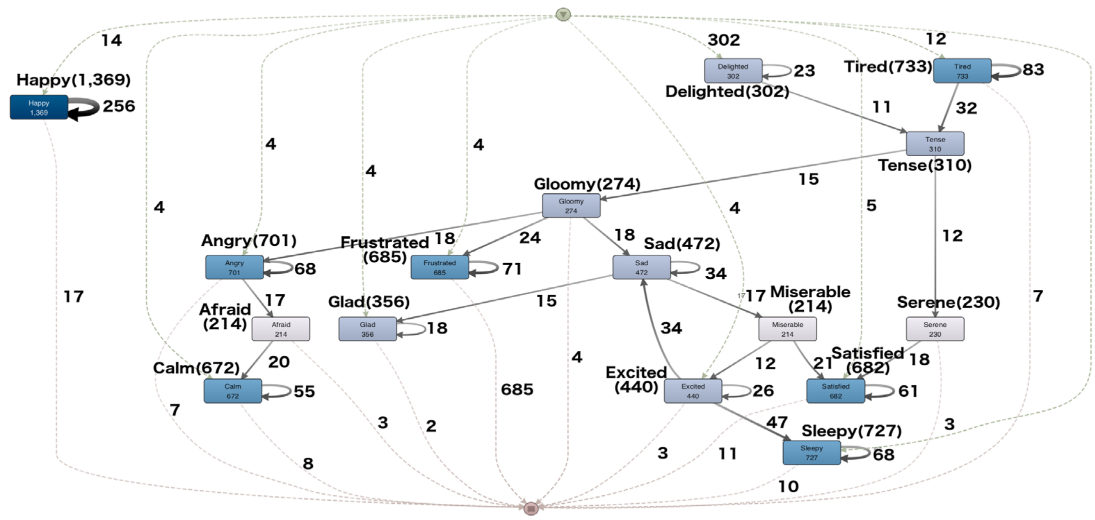


Fig. 6: A screenshot of the generated Disco Fuzzy Miner process model for 100% of Activities and 50% of Paths, based on the PAM emotional data during StudentLife project

By assessing the resulting Disco Fuzzy Miner emotional model shown in Fig. 6 and after investigating the relationship amongst the emotional indicators, it is understood that the activities “Afraid” and “Calm” in most of the cases occurred after “Angry”. In the same way, the activities “Sleepy” and “Sad” in most of the cases are occurred after “Excited”. Accordingly, the Disco Fuzzy Miner process mining technique can be used to provide qualitative information in regard to the emotional state of the students during the StudentLife project.

4.2 “Severity” of Depression-Related Emotions based on 5 Clusters

The emergence and manifestation of emotions are specific to every different individual [22]. Obviously, every person has a different emotional pattern depending on the environment and the specific circumstances of the individual. In this research, the target students were exposed to the same environment at the same time. This approach enabled us to determine whether different groups of depression have a similar (or dissimilar emotional patterns) or not. Table 1 shows the number of students suffering from depression (with its extent of severity) in five groups. The scores were obtained and calculated based on the PHQ-9 survey test (questionnaire) which were distributed and collected both before and after the commencement of the StudentLife project. The reason why in this work the depression model was divided into 5 scales (based on the PHQ-9) was due to the fact that a similar approach[4], exactly compatible with the works previously done by [23] and [24], which were aimed to be followed

[4] For more information about the PHQ-9 and its 5 models please check out the link below:

and practiced. As shown in Table 1, according to the PHQ-9 survey test (questionnaire), which is a 9-item depression module, the total results of the ratings (containing both pre and post survey) are divided into five depression-associated categories: (1) Students with minimal degree of depression severity. (2) Students with minor degree of depression severity. (3) Students with moderate degree of depression severity. (4) Students with moderately severe degree of depression severity. (5) Students with severe degree of depression severity.

Table 1: PHQ-9 depression scale interpretation (pre and post survey) [8], [23-24]

Depression severity	**minimal**	**minor**	**moderate**	**Moderately severe**	**severe**
Total Score	1-4	5-9	10-14	15-19	20-27
Number of students (pre-test)	19	17	6	1	1
Number of students (post-test)	15	12	3	2	2

As Fig. 7 shows, the response rate to the PAM questionnaire survey (based on the selected image) has been reduced by one passing of time as the total number of dots has been dropped at the end of each cluster. In addition, Fig. 7 illustrates a comparison of the response volume in terms of the duration of time for all of the randomly selected five students from each severity-of-depression categories. As illustrated in Fig. 7, the response rate at the beginning of the PAM questionnaire survey decreases over time. This is compatible with the EMA results previously obtained from the StudentLife research project.

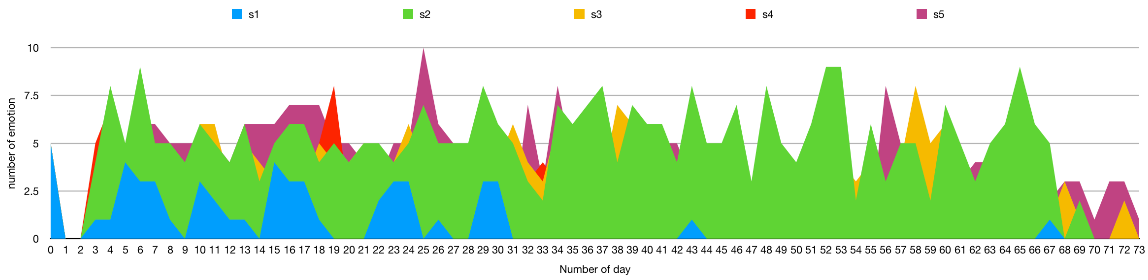
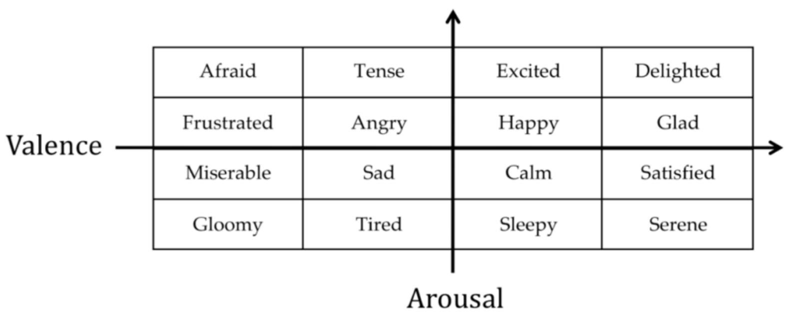


Fig. 7: Comparison Relative frequency per day for 5 students, each individual randomly selected and chosen from each of the 5 categories of the PHQ-9 matrix/model according to the "severity type" of the depression-related emotions

4.3 Categorization or Grouping of the "16 Emotional Indicators" into 4 Quadrants

In order to better study and investigate the students' emotional trends (and emotional behavior) based on the Russell's Circumplex Model of Affect Grid, in this study all of Russell's 16 emotional values were classified and divided into four main groups in a similar way previously conducted by Watson & Tellegen (1985) [25]. The rationale behind such categorization was to examine the tendency of emotions in terms of "High Positive", "Low Positive", "High Negative" and "Low Negative" moods (see Fig. 8, Fig. 9 and Table 2).



http://www.lphi.org/LPHIadmin/uploads/.PHQ-9-Review-Kroenke-63754.PDF

Fig. 8: Categorization or Grouping of the "16 Emotional Indicators" into four main clusters/quardants according to the Russell's Circumplex Model of Affect Grid [26]

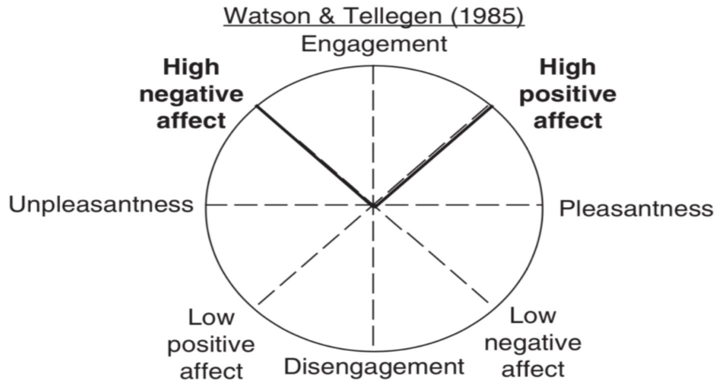


Fig. 9: Categorization and grouping of the Affect Circumplex Model into four main moods based on the "Intensity" of their impact or influence [25]

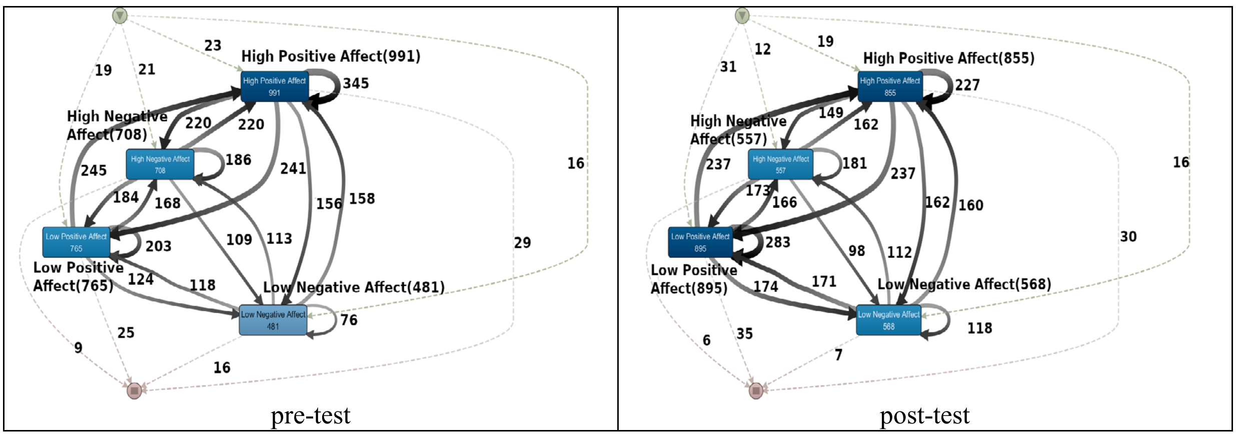
Table 2 provides a better understanding in regard to the way such grouping of the positive or negative emotional tendencies (over16 emotional indicators/values) was done and applied as follows:

- High Negative Affect (High-NA): Frustrated, Afraid, Angry and Tense.
- High Positive Affect (High-PA): Happy, Excited, Glad and Delighted.
- Low Negative Affect (Low-PA): Gloomy, Miserable, Tired and Sad.
- Low Positive Affect (Low-NA): Sleepy, Calm, Serene and Satisfied.

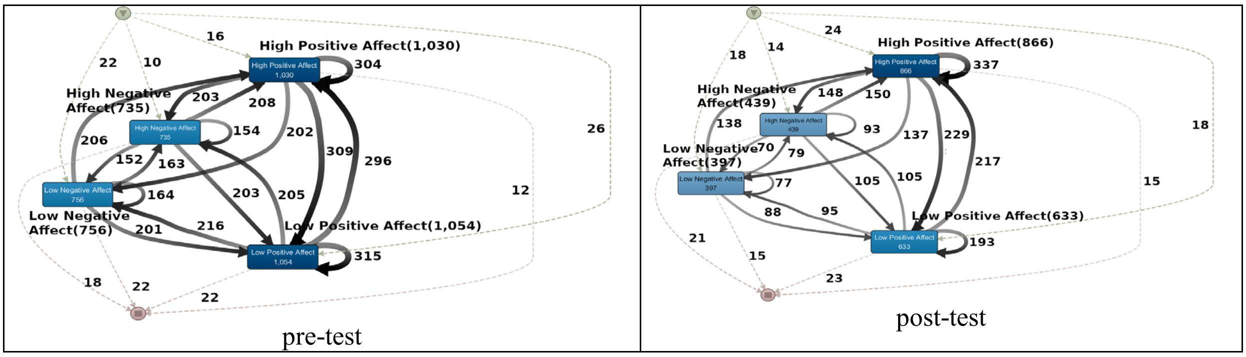
Table 2: Grouping and scoring the "Affect Circumplex Model" constructs into four main quadrants based on the intensity of their influence/affect in addition to the allocated rates

Emotion Type (allocated scores)

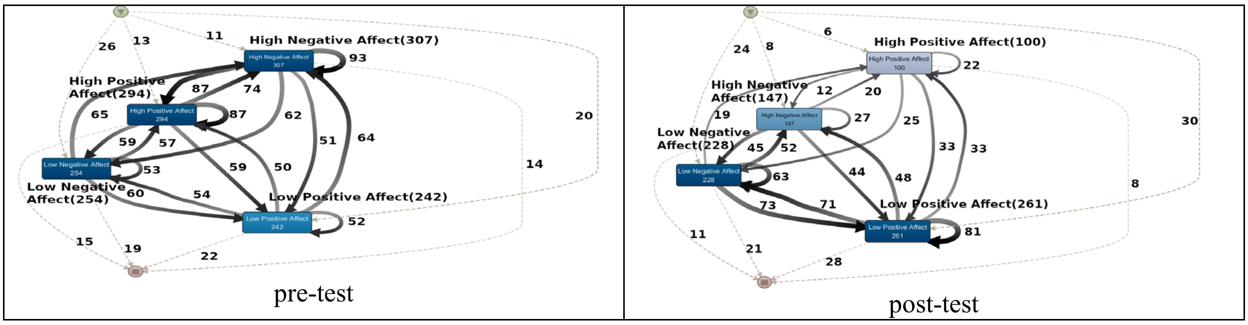
High Negative Affect		**High Positive Affect**	
Afraid (-8)	Tense (-7)	Excited (+8)	Delighted (+7)
Frustrated (-6)	Angry (-5)	Happy (+6)	Glad (+5)
Low Negative Affect		**Low Positive Affect**	
Miserable (-4)	Sad (-3)	Calm (+4)	Satisfied (+3)
Gloomy (-2)	Tired (-1)	Sleepy (+2)	Serene (+1)



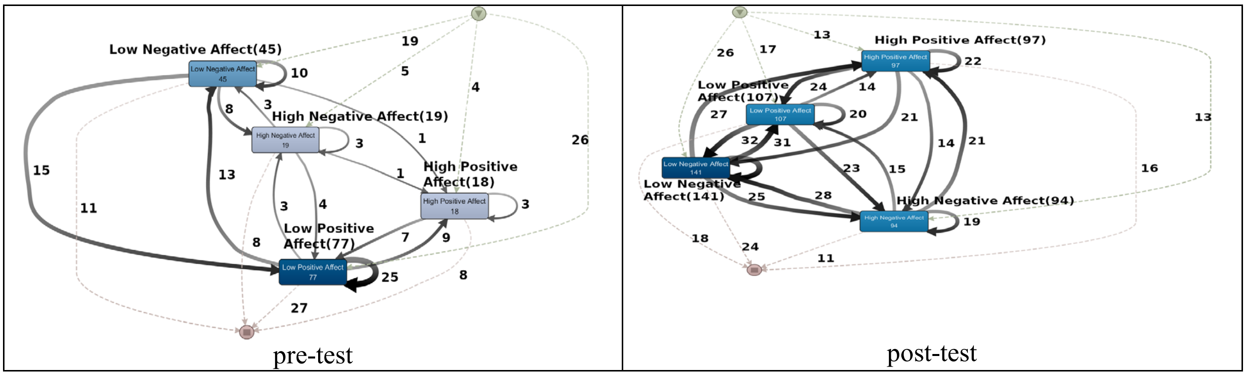
(a) Illustrations of the depression-related emotions with "Minimal Degree of Severity"



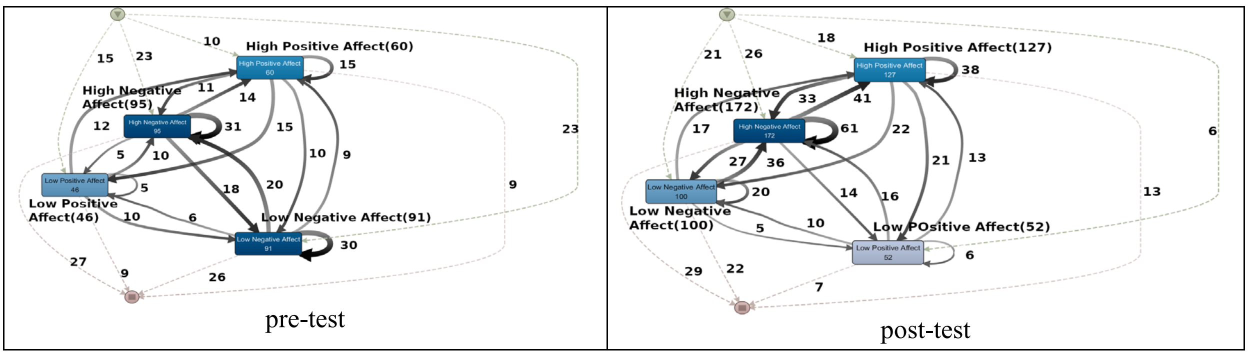
(b) Illustrations of the depression-related emotions with "Minor Degree of Severity"



(c) Illustrations of the depression-related emotions with "Moderate Degree of Severity"



(d) Illustrations of the depression-related emotions with "Moderately High Degree of Severity"



(e) Illustrations of the depression-related emotions with "Very High Degree of Severity"

Fig. 10: A frequency-based comparison of the "Emotional Map Affect Grid" constructs in terms of four categories (based on the severity type of the depression-related emotions) through Disco process mining tool and via "Fuzzy Miner" algorithm for both pre-test and post-test scenarios [27-30]

Fig. 10 (a-e) show a holistic illustration/comparison of the "Emotional Map Affect Grid" in terms of 4 main categories based on the severity type of the depression-related emotions through process mining "Disco Fuzzy Miner graphs/models" for both StudentLife's pre-test and post-test experiments.

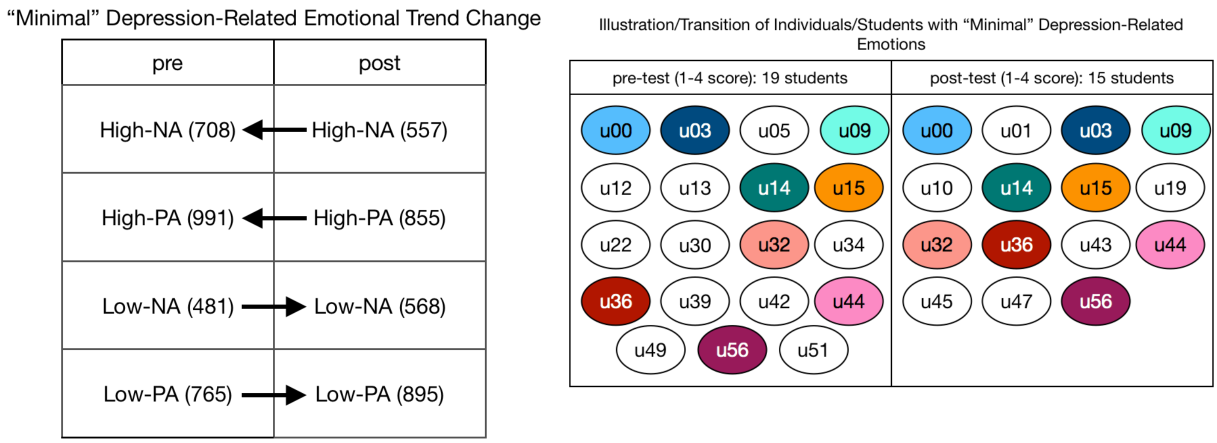


Fig. 11: *(left)* An illustration of the " "Directional Changes" of the students who manifested depression-related emotions with "Minimal Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios. *(right)* An illustration of the "Transition Shift" or "Displacement Changes" of the students who manifested depression-related emotions with "Minimal Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios.

Fig. 10 (a) and Fig. 11 (left) show two illustrations of the "frequency-based" and "directional change" views of the "minimal depression-related emotions" for constructs containing Positive and Negative emotions/moods based on High and Low extents of severity, respectively. Within the "pre-test" stage, the extent of Low-PA mood was 765 while the extent of High-PA was 991. Within the "post-test" stage, the extent of Low-PA mood was 895 while the extent of High-PA was 855. All of the frequency numbers in Fig. 11 (left) to Fig. 15 (left) were obtained from the resulting Fuzzy Miner graphs (within a frequency-based approach) earlier represented in Fig. 10 (a-e). By further contemplating on the "frequency-based" results of "minimal depression-related emotions" from low using Disco Fuzzy Miner graphs/models, it is obvious that a significant change from Low-NA to High-PA had occurred during the "pre-test" stage. On the other hand, a significant change from High-PA to Low-PA had occurred during the "post-test" stage.

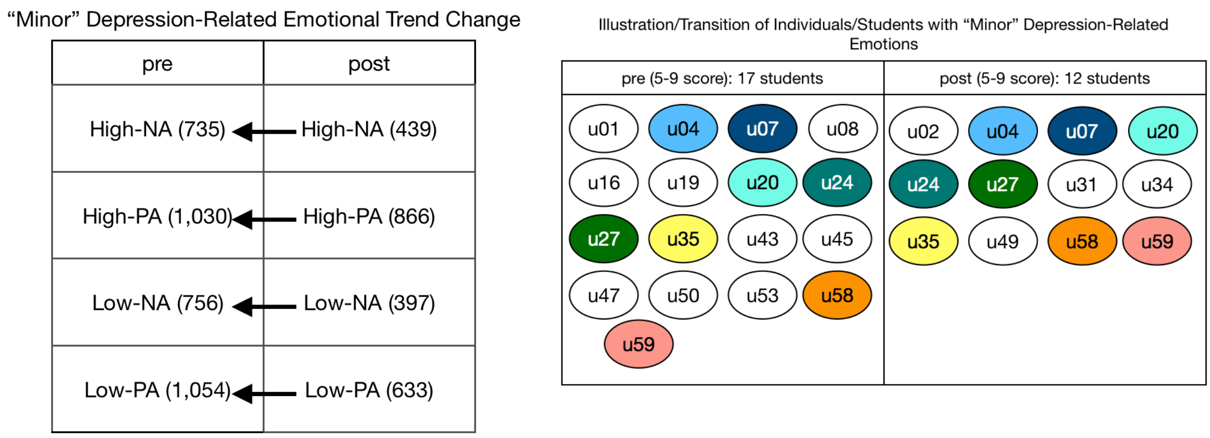


Fig. 12: *(left)* An illustration of the " "Directional Changes" of the students who manifested depression-related emotions with "Minor Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios. *(right)* An illustration of the "Transition Shift" or "Displacement Changes" of

the students who manifested depression-related emotions with “Minor Degree of Severity”, based on the frequency or number of times they were repeated during the pre-test and post-test scenarios.

By studying the results of the “Transition Trend Shift” of the severity of the depression-related emotions (or simply called “Directional Displacement Changes”) shown in Fig. 11 (right) to Fig. 15 (right), it was understood that 11 students (i.e., u05, u08, u12, u13, u22, u30, u39, u42, u46, u50, u51) did not contribute/participate in the post-test exam. Moreover, as shown in Fig. 11 (right), the severity type of the depression-related emotions for 9 students (i.e., u00, u03, u09, u14, u15, u32, u36, u44, u56) remained “minimal” and completely unchanged during the pre-test and post-test activities. The “severity” of the depression-related emotions for 2 students (i.e., u34, u49) changed/increased from “minimal” degree of severity during the pre-test activity to “minor” degree of severity during the post-test activity.

Fig. 10 (b) and Fig. 12 (left) show two illustrations of the “frequency-based” and “directional change” views of the “minor depression-related emotions” for constructs containing Positive and Negative emotions/moods based on High and Low extents of severity, respectively. Within the “pre-test” stage, the extent of Low-PA mood was 1054 while the extent of High-PA was 1030. Within the “post-test” stage, the extent of Low-PA mood was 633 while the extent of High-PA was 866. All of the frequency numbers in Fig. 12 (left) were obtained from the resulting Fuzzy Miner graphs (through a frequency-based approach) earlier represented in Fig. 10 (b). By further contemplating on the “frequency-based” results of “minor depression-related emotions” from low using Disco Fuzzy Miner graphs/models, it is obvious that a significant change from High-PA to Low-NA had occurred during the “pre-test” stage. On the other hand, a significant change from Low-PA to High-PA had occurred during the “post-test” stage.

By studying the results of the “Transition Trend Shift” of the severity of the depression-related emotions (or “Directional Displacement Changes”) shown in Fig. 12 (right), it was understood that the severity type of the depression-related emotions for 6 students (i.e., u20, u24, u27, u35, u58, u59) remained “minor” and completely unchanged during the pre-test and post-test activities. However, the “severity” of the depression-related emotions for 5 students (i.e., u01, u19, u43, u45, u47) changed/decreased from “minor” degree of severity during the pre-test activity to “minimal” degree of severity during the post-test activity. Quite opposite, the “severity” of the depression-related emotions for 2 students (i.e., u16, u53) changed/increased from “minor” degree of severity during the pre-test activity to “moderate” degree of severity during the post-test activity.

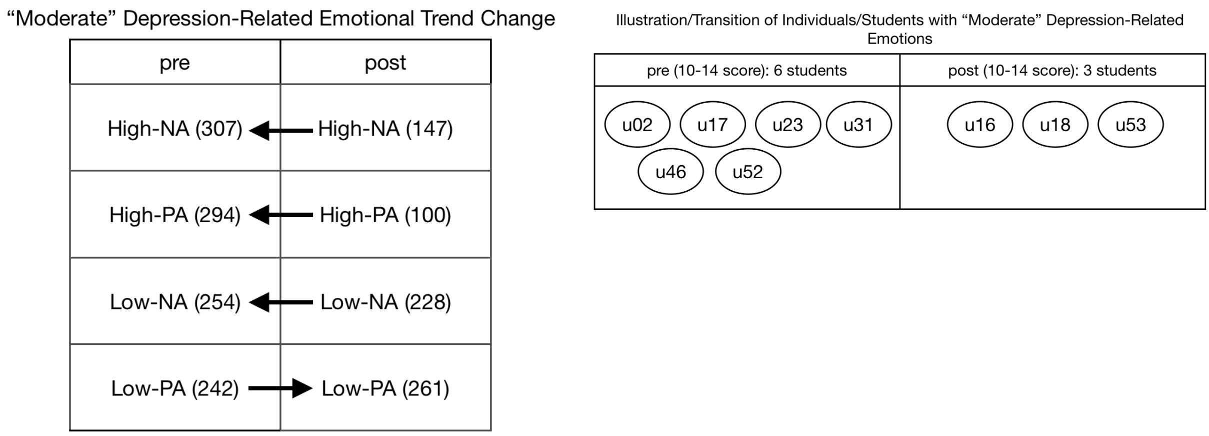


Fig. 13: *(left)* An illustration of the “ “Directional Changes” of the students who manifested depression-related emotions with “Moderate/Average Degree of Severity”, based on the frequency or number of times they were repeated during the pre-test and post-test scenarios. *(right)* An illustration of the “Transition Shift” or “Displacement Changes” of the students who manifested depression-related emotions with “Moderate/Average Degree of Severity”, based on the frequency or number of times they were repeated during the pre-test and post-test scenarios.

Fig. 10 (c) and Fig. 13 (left) show two illustrations of the “frequency-based” and “directional change” views of the “moderate depression-related emotions” for constructs containing Positive and Negative emotions/moods based on High and Low extents of severity, respectively. Within the “pre-test” stage, the extent of High-NA mood was 307 while the extent of High-PA was 294. Within the “post-test” stage, the extent of Low-PA mood was 261 while the extent of Low-NA was 228. All of the frequency numbers in Fig. 13 (left) were obtained from the resulting Fuzzy Miner graphs (within a frequency-based approach) earlier represented in Fig. 10 (c). By further contemplating on the “frequency-based” results of “moderate depression-related emotions” from low using Disco Fuzzy Miner graphs/models, it is obvious that a significant change from Low-NA to High-PA had occurred during the "pre-test" stage. On the other hand, a significant change from High-PA to High-NA had occurred during the "post-test" stage.

By studying the results of the "Transition Trend Shift" of the severity of the depression-related emotions (or "Directional Displacement Changes") shown in Fig. 13 (right), it was found out that the severity type of the depression-related emotions for 2 students (i.e., u02, u31) changed/decreased from "moderate" degree of severity during the pre-test activity to "minor" degree of severity during the post-test activity. Quite opposite, the "severity" of the depression-related emotions for 2 students (i.e., u17, u52) changed/increased from "moderate" degree of severity during the pre-test activity to "moderately severe" during the post-test activity. Similarly, the "severity" of the depression-related emotions for 1 student (i.e., u23) changed/increased from "moderate" degree of severity during the pre-test activity to "completely severe" level during the post-test activity.

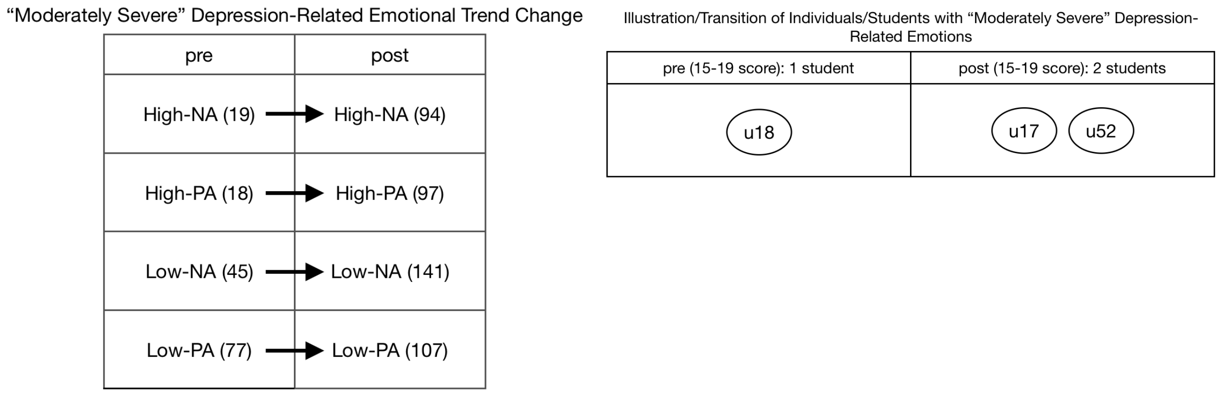


Fig. 14: *(left)* An illustration of the " "Directional Changes" of the students who manifested depression-related emotions with "Moderately High Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios. *(right)* An illustration of the "Transition Shift" or "Displacement Changes" of the students who manifested depression-related emotions with "Moderately High Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios.

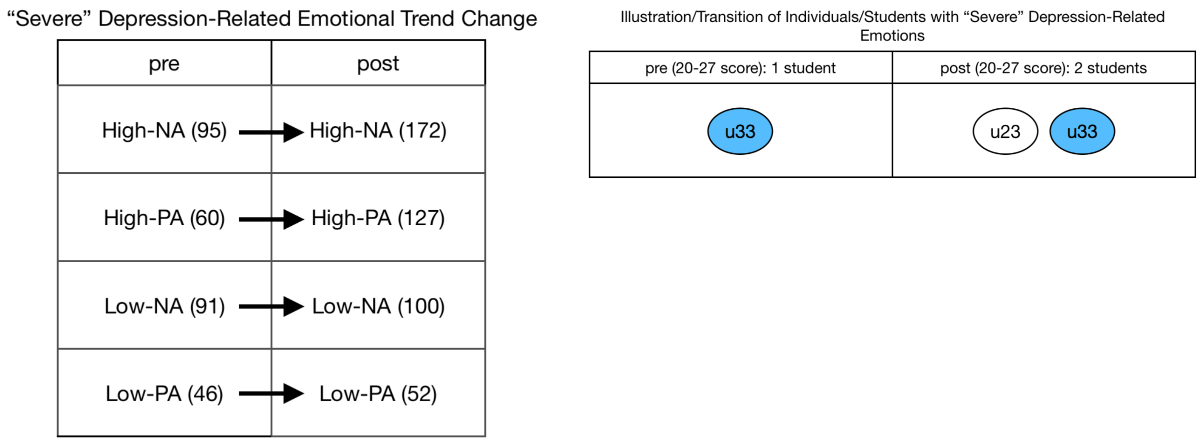


Fig. 15: *(left)* An illustration of the " "Directional Changes" of the students who manifested depression-related emotions with "Extremely High Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios. *(right)* An illustration of the "Transition Shift" or "Displacement Changes" of the students who manifested depression-related emotions with "Extremely High Degree of Severity", based on the frequency or number of times they were repeated during the pre-test and post-test scenarios.

Fig. 10 (d) and Fig. 14 (left) show two illustrations of the "frequency-based" and "directional change" views of the "moderately severe depression-related emotions" for constructs containing Positive and Negative emotions/moods based on High and Low extents of severity, respectively. Within the "pre-test" stage, the extent of Low-PA mood was 77 while the extent of Low-NA was 45. Within the "post-test" stage, the extent of Low-PA mood was 107 while the extent of Low-NA was 141. All of the frequency numbers in Fig. 14 (left) were obtained from the resulting Fuzzy Miner graphs (within a frequency-based approach) earlier represented in Fig. 11 (d). By further analysis of the "frequency-based" results of "moderately severe depression-related emotions" from low using Disco Fuzzy Miner graphs/models, it is obvious that a significant change from Low-PA to Low-NA had occurred during the "pre-test" stage. On the other hand, a significant change from Low-NA to Low-PA had occurred during the "post-test" stage. Furthermore, by studying the results of the "Transition Trend Shift" of the severity of the depression-related emotions (or "Directional Displacement Changes") shown in Fig. 14 (right), it was found that the severity type of the depression-related emotions for 1 student (i.e., u18) changed/decreased from "moderately severe" during the pre-test activity to "moderate" level of severity during the post-test activity.

Fig. 10 (e) and Fig. 15 (left) show two illustrations of the "frequency-based" and "directional change" views of the "severe depression-related emotions" for constructs containing Positive and Negative emotions/moods based on High and Low extents of severity, respectively. Within the "pre-test" stage, the extent of High-NA mood was 95 while the extent of Low-NA was 91. Within the "post-test" stage, the extent of High-NA mood was 172 while the extent of High-PA was 127. All of the frequency numbers in Fig. 15 (left) were obtained from the resulting Fuzzy Miner graphs (within a frequency-based approach) earlier represented in Fig. 10 (e). By further analysis of the "frequency-based" results of "severe depression-related emotions" from low using Disco Fuzzy Miner graphs/models, it is obvious that a significant change from Low-PA to High-NA had occurred during the "pre-test" stage. On the other hand, a significant change from High-PA to High-NA had occurred during the "post-test" stage. Moreover, after further investigation of the results of the "Transition Trend Shift" of the severity of the depression-related emotions (or simply called "Directional Displacement Changes") shown in Fig. 15 (right), it was understood that the severity type of the depression-related emotions for 1 student (i.e., u33) remained "completely severe" and unchanged during the pre-test and post-test activities.

5.0 CONCLUSION AND FUTURE WORK

Process Mining is a new process management field that has recently received much attention within the last few years. Process mining and process modeling has the potential to become a useful approach for understanding and analyzing the behavior of people; especially when dealing with a large amount of datasets and information. Fortunately nowadays more and more data (so-called event logs) in regard to the 'activity processes' is recorded by a variety of different Information Systems. However, one of the most interesting trends to study (and apply process mining techniques for process visualization purposes) is Smartphones, which are almost available to everyone at any time. In this research, a novel approach to investigate and analyze students' emotions based on the StudentLife data was proposed and implemented. The main idea was based on the assessment of the mood of the Smartphone users through two process mining techniques; namely, "Fuzzy Miner" and "Dotted Chart Analysis". Accordingly, the current work focuses on the application of the process mining tools and techniques with the purpose of discovering and analyzing emotional patterns of the sample group of students participating in the StudentLife project, where the data sources were freely and openly published/accessible to public for research purposes [8]. The collected data included a set of emotional data derived from the PAM (Photographic Affection Meter) application, which interrelated and identified the students' emotions based on the chosen photos. Within the data preparation and pre-processing phase, the Python programming language was used in order to convert the format of the initially collected (raw) files/data from JSON to CSV (see Fig. 1). After the data cleansing process of the PAM data, the "Circumplex's Affect Grid" model was used in order to specify and identify the students' emotions rooted in 16 emotional indicators (i.e., Afraid, Tense, Frustrated, Angry, Miserable, Sad, Gloomy, Tired, Excited, Delighted, Happy, Glad, Calm, Satisfied, Sleepy, Serene) and categorized into four main quadrants (i.e., depression-related emotions with "minimal", "minor", "moderate", "moderately severe" and "very severe" degrees/levels of severity) in terms of Valence and Arousal coordinates (see Fig. 9, Fig. 10, Table 1 and Table 2) [24]. In order to streamline the process modeling (and simulation) of the students' 16 emotional indicators in a "frequency-based" approach (i.e., number of times the process instances were executed and performed within the data), the Disco Fluxicon, which is a popular process mining tool/platform and works based on the Fuzzy Miner algorithm, was used. Moreover, in order to better visualize the type of students' emotion within a specific span of time, the Dotted Chart Analysis technique, supported by ProM process mining tool/platform, was applied as well (see Fig. 2 and Fig. 3). Such techniques enabled us to better demonstrate an overview of the emotional processes in addition to the emotional trends. The resulting Fuzzy Miner graphs/models could clearly indicate the frequency, links and path changes of emotions occurring in each session/period of the semester. Although the ProM Fuzzy Miner technique was capable of indicating the extent of relationships amongst all of the pre-defined 16 emotional indicators occurring within a 64-day project experiment, the "spaghetti-like" nature of the resulting process model (see Fig. 5), which was associated with less structured processes, made it difficult to comprehend, interpret and analyze the resulting dependencies between the tasks, making it too difficult to interpret the data, or gain some insights into the PAM emotional event logs. To deal with this problem, a threshold simplification cutoff (with 100% Activities and 50% of Paths) was implemented on the StudentLife data. Using this approach made it possible to reduce and simplify the number of activities or the relationship amongst the activities and paths (see Fig. 6). After applying Disco Fuzzy Miner algorithm on the PAM StudentLife data (including 16 emotional indicators), the results showed that the emotions “Happy (16.33%)”, “Tired (8.75%)”, “Sleepy (8.67%)” and “Angry (8.36%)” allocated the highest relative frequencies, respectively (see Fig. 4).

Moreover, by following a Path Reduction method, it was more convenient to investigate the emotional relationships amongst the indicators (values) in a more specific and understandable manner according to four depression-related severity of emotions. Accordingly, a holistic comparison of the “Emotional Map Affect Grid” for all of the 43 students participated in this study in terms of 4 categories via process mining Disco Fuzzy Miner graphs/models, for both pre-test and post-test scenarios, was represented and illustrated, as shown in Fig. 10 (a-e).
As shown in Fig. 11 (left), after analysis of the generated Fuzzy Miner graphs (i.e., within a frequency-based approach) related to the depression-related emotions accompanied with “Minimal” degree/level of severity”, the following “Trend Shift” or “Directional Changes” were observed and identified as follows:

- The extent of the High Negative Affections (708) during the pre-test was greater than the extent of Negative Affections (557) during the post-test
- The extent of the High Positive Affections (991) during the pre-test was greater than the extent of Positive Affections (855) during the post-test
- The extent of the Low Negative Affections (481) during the pre-test was less than the extent of Low Negative Affections (568) during the post-test
- The extent of the Low Positive Affections (765) during the pre-test was less than the extent of Low Positive Affections (895) during the post-test

As shown in Fig. 12 (left), after analysis of the generated Fuzzy Miner graphs (i.e., within a frequency-based approach) related to the depression-related emotions accompanied with “Minor” degree/level of severity”, the following “Trend Shift” or “Directional Changes” were observed and identified as follows:

- The extent of the High Negative Affections (735) during the pre-test was greater than the extent of Negative Affections (439) during the post-test
- The extent of the High Positive Affections (1,030) during the pre-test was greater than the extent of Positive Affections (866) during the post-test
- The extent of the Low Negative Affections (756) during the pre-test was greater than the extent of Low Negative Affections (397) during the post-test
- The extent of the Low Positive Affections (1,054) during the pre-test was greater than the extent of Low Positive Affections (633) during the post-test

As shown in Fig. 13 (left), after analysis of the generated Fuzzy Miner graphs (i.e., within a frequency-based approach) related to the depression-related emotions accompanied with “Moderate” degree/level of severity”, the following “Trend Shift” or “Directional Changes” were observed and identified as follows:

- The extent of the High Negative Affections (307) during the pre-test was greater than the extent of Negative Affections (147) during the post-test
- The extent of the High Positive Affections (294) during the pre-test was greater than the extent of Positive Affections (100) during the post-test
- The extent of the Low Negative Affections (254) during the pre-test greater than the extent of Low Negative Affections (228) during the post-test
- The extent of the Low Positive Affections (242) during the pre-test was less than the extent of Low Positive Affections (261) during the post-test

As shown in Fig. 14 (left), after analysis of the generated Fuzzy Miner graphs (i.e., within a frequency-based approach) related to the depression-related emotions accompanied with “Moderately Severe” degree/level of severity”, the following “Trend Shift” or “Directional Changes” were observed and identified as follows:

- The extent of the High Negative Affections (19) during the pre-test was less than the extent of Negative Affections (94) during the post-test
- The extent of the High Positive Affections (18) during the pre-test was less than the extent of Positive Affections (97) during the post-test
- The extent of the Low Negative Affections (45) during the pre-test was less than the extent of Low Negative Affections (141) during the post-test

- The extent of the Low Positive Affections (77) during the pre-test was less than the extent of Low Positive Affections (107) during the post-test

As shown in Fig. 15 (left), after analysis of the generated Fuzzy Miner graphs (i.e., within a frequency-based approach) related to the depression-related emotions accompanied with “Very Severe” degree/level of severity”, the following “Trend Shift” or “Directional Changes” were observed and identified as follows:

- The extent of the High Negative Affections (95) during the pre-test was less than the extent of Negative Affections (172) during the post-test
- The extent of the High Positive Affections (60) during the pre-test was less than the extent of Positive Affections (127) during the post-test
- The extent of the Low Negative Affections (91) during the pre-test was less than the extent of Low Negative Affections (100) during the post-test
- The extent of the Low Positive Affections (46) during the pre-test was less than the extent of Low Positive Affections (52) during the post-test

In order to provide the reader with a better understanding about the benefits and advantages of the proposed technique in addition to the above-stated results, an example is given. Let us assume that an instructor (or a researcher) is interested in investigating the “Emotional Behavior” of a randomly chosen student (ID Number of U34 in this case). Fig. 16 shows an obvious “Transition Shift” of the depression-related emotions for “Student U34” between the pre-test and post-test activities. “Student U34” manifested depression-related emotions with “Minimal (Very Low) Degree/Level of Severity” during the pre-test activity, while the level of depression was slightly increased to “Minor (Low)” after the post-test activity. In other words, the level of “severity of depression-related emotions” was boosted and increased during the assigned task.

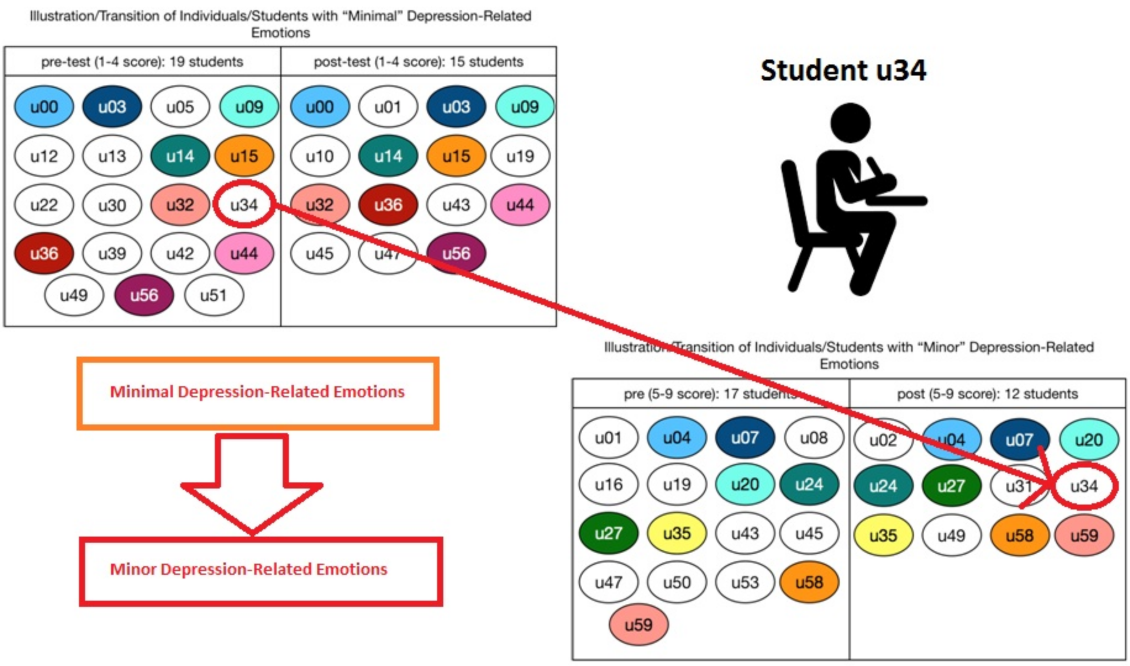


Fig. 16: An illustration of the depression-related “Transition Shift” or “Displacement Trend Changes” of the emotions for “Student U34” from “Minimal Level” (during the pre-test activity) to “Minor Level” (during the post-test activity)

Although, being aware of the severity of the depression-related emotions during the pre and post tests (shown in Fig. 16) can be interesting, the instructor might be interested in knowing more details about fluctuations of the emotions throughout the course (and during the period of time when the StudentLife project had been running) as well. To do this, a new representation of the “Affect Circumplex Model” constructs and their allocated scores/weights in terms of a two dimensional coordinate space based on the “Day Numbers” and “Allocated Emotion Scores” was applied and used. By contemplating on the “Fluctuations” of the emotional Affect Circumplex indicators shown in Fig. 17, the instructor can gain more insights about the fact that on 14th day of the experiment, the “Student U34” underwent and experienced the highest level of Negative Affects (i.e., depress-related emotions was with a very high degree of severity). Similarly, during the days 6th, 15th and 19th, the "severity level" of the stress-related negative emotions was quite high.

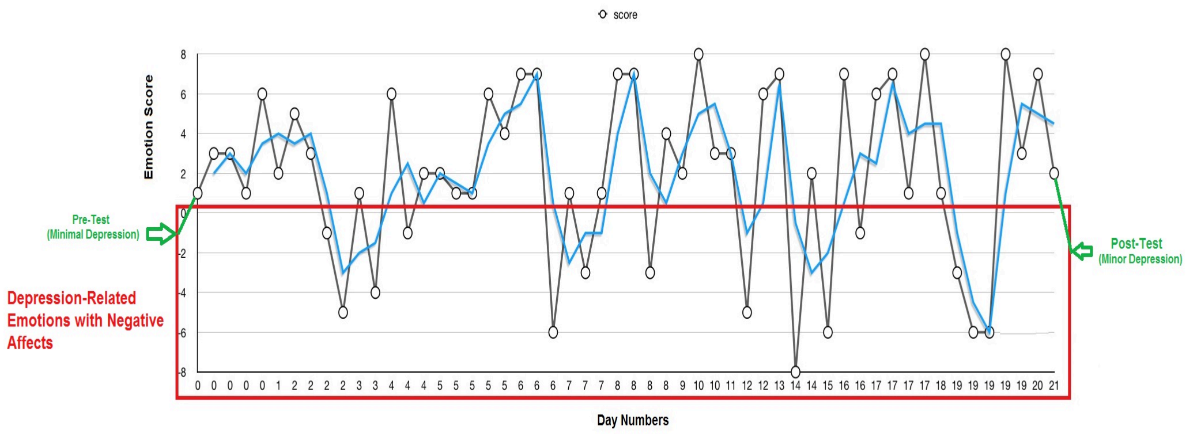


Fig. 17: An illustration of the emotional "Fluctuations" during the ongoing StudentLife project based on the "Affect Circumplex Model" constructs and their allocated scores for "Student U34"

Furthermore, not only the emotional fluctuations but the "type" of the emotional indicators (over a specific span of time) also can be discovered and studied by the instructor. To do this, the process mining "Dotted Chart Analysis" technique (supported by ProM process mining tool/platform) can be used in order to identify the main causes of the type of the (both positive and negative) emotional affects throughout the StudentLife project/experiment as shown in Fig. 18.

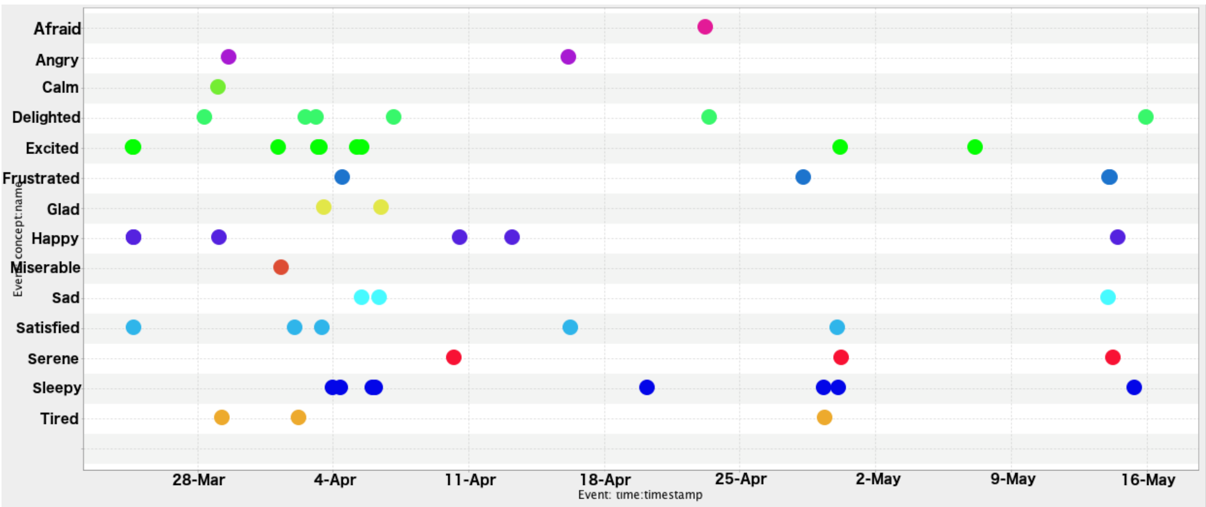


Fig. 18: A dotted illustration of the "types" of the emotional affects for "Student U34", including both positive and negative indicators/constructs, during the ongoing StudentLife project based on the "Affect Circumplex Model" and by using the "Dotted Chart Analysis" technique supported by ProM process mining tool/platform

And eventually, not only "transition shifts", "fluctuations" and "types" of the (both positive and negative) emotional indicators, but the "frequency" of the emotional variables/constructs also can be investigated by the instructor through a frequency-based Fuzzy Miner approach and via Disco Fluxicon process mining tool/platform, as shown in Fig. 19.

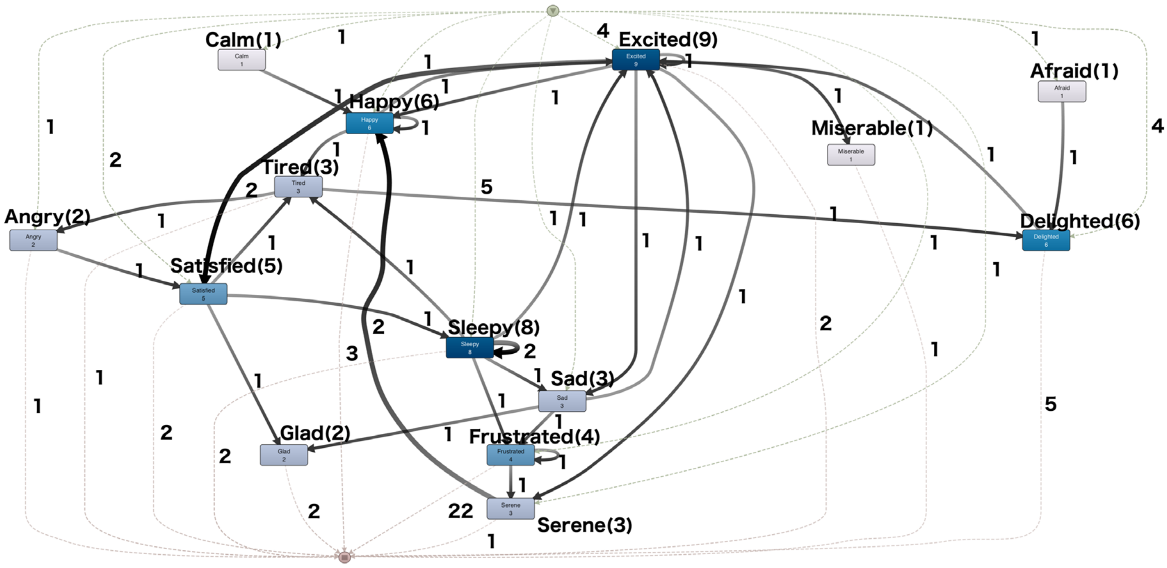


Fig. 19: A “frequency-based” illustration of the emotional affects for “Student U34”, including both positive and negative indicators/constructs, during the ongoing StudentLife project based on the “Affect Circumplex Model” and by using the “Fuzzy Miner” technique/algorithm supported by Disco Fluxicon process mining tool/platform

To conclude and in order to compare the techniques proposed (and applied) in this study with other works or (techniques) previously done (or applied) in this field, both of the “practical” and “theoretical” contributions of this research can be briefly expressed and formulated as the following: (i) Practically, for the first time, “Fuzzy Miner (via Disco Fluxicon)” and “Dotted Chart Analysis (via ProM)” process mining techniques and tools were applied and implemented on the StudentLife data. The main advantages of the two adopted process mining techniques were the objective and quick diagnosis of emotional depression-related changes and issues. The novel approach was capable of significantly facilitating the simulation process of the “depression-related emotions” amongst the students, leading to a better understanding and improvement of mental health and academic performance, by extracting the necessary information out of the existing StudentLife data. (ii) Theoretically, for the first time, both “Frequency-Based *(or Performance Analysis)*” and “Directional Displacement Change *(or Transition Trend Shift)*” graphical simulations/illustrations of the both “extent” and “severity type” of the “depression-related emotions”, respectively, during pre-test and post-test scenarios, were used, analyzed and represented.

Consequently, by discovering, visualizing, simulating and understanding the “emotional behavior” and “emotional tendencies” of a student, instructors can increase the level of their awareness toward the “emotional contributions” of the students during an assigned task leading to improvement of the coaching or teaching style in such ways that can increase their students’ mental health and academic performance. From the student point of view, the proposed approach discussed in this study can help the students to better manage their emotions during the ongoing projects or assigned tasks. Moreover, such data can significantly boost the level of their self-concept (i.e., mirroring) toward their weak (and strong) emotional points. However, this research provides groundwork for further studies. In the future, we are going to study and investigate the relationship between the “emotional behavior” and “depression-related indicators” of the students with their final GPA or performance analytics [31] during the assigned task.

REFERENCES

[1] State of Mental Health in America Report Overview Historical Data, *http://www.mentalhealthamerica.net/issues/2017-state-mental-health-america-report-overview-historical-data, 2017.*

[2] J. P. Pollak, P. Adams, and G. Gay, "PAM: A Photographic Affect Meter for Frequent, In Situ Measurement of Affect". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada*, 07–12 May 2011, pp. 725-734.

[3] Van der Aalst, W., *Process Mining: Discovery, Conformance and Enhancement of Business Processes* Springer., 2011.

[4] Fluxicon Process Laboratories*, https://fluxicon.com/disco/, 2012.*

[5] Eindhoven University of Technology*, http://www.promtools.org/doku.php, 2010.*

[6] M. Bozkaya, J. Gabriels, J. M. van der Werf , "Process Diagnostics:a Method Based on Process Mining", in *International Conference on Information, Process, and Knowledge Management, eKNOW 2009, Cancun, Mexico,* 1-7 February 2009.

[7] R. Wang, G. Harari†, P. Hao, X. Zhou, and Andrew T. Campbell, "SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students", in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp*, 2015, pp. 295–306.

[8] Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell, "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones", in *Proceedings of the ACM Conference on Ubiquitous Computing, UbiComp*, 2014, pp. 3–14.

[9] Automation and Process Mining: A Powerful Combination, *https://fluxicon.com/blog/2016/07/automation-platforms-and-process-mining-a-powerful-combination/, 26 July 2016.*

[10] Easterbrooke J A., "The effect of emotion on cue utilization and the organization of behavior." *Psychological Review*, Vol. 66, 1959, pp. 187-201.

[11] Russell, J.A., Weiss, A., Mendelsohn, G.A.,"Affect Grid: A Single-Item Scale of Pleasure and Arousal", *Journal of Personality and Social Psychology*, Vol. 57 No. 3, 1989, pp. 493-502.

[12] D. Watson, L. A. Clark, and A. Tellegen, "Development and Validation of Brief Measures of Positive and Negative Affect:The PANAS Scales", *Journal of Personality and Social Psychology*, Vol. 54 No. 6, 1988, pp. 1063-1070.

[13] Karl Halvor Teigen, Yerkes-Dodson, "A Law for all Seasons", *Theory & Psychology*, SAGE, Vol. 4 No. 4. November 1994, pp. 525-547.

[14] Reisenzein, R., "Pleasure-arousal theory and the intensity of emotions", *Journal of Personality and Social Psychology*, Vol. 67 No. 3, 1994, pp. 525-253.

[15] Watson D., Wiese D., Vaidya J., Tellegen A., "The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence" *Journal of Personality and Social psychology*, Vol. 76 No. 5, 1999, pp. 820-838.

[16] Plutchik R., "The Nature of Emotions", *American Scientist* Vol. 89, 2001, pp. 344-350.

[17] Yang Shen, Shuchen Li, Ling Zheng, Xiaodong Ren, Xiaolong Cheng, Wuhan Univ., Wuhan, China, "Emotion Mining Research on Micro-Blog", in *Proceeedings of 1st IEEE Symposium on Web Society, Lan Zhou, China*, 2009, pp. 71-75.

[18] Seyyed Abed Hosseini, Mohammad Ali Khalilzadeh, Mashhad Branch, Biomed. Eng. Dept., Islamic Azad Univ., Mashhad, Iran, "Emotional stress recognition system using EEG and psychophysiological signals:Using new labelling process of EEG signals in emotional stress state", in *2010 International Conference on Biomedical Engineering and Computer Science, Wuhan, China*, April 2010.

[19] S. Haim, R. Wang, S. E. Lord, L. Loeb, X. Zhou, A.T.Campbell, Dartmouth College, "The Mobile Stress Meter: A New Way to Measure Stress Using Images", *ACM Workshop on Mobile Systems for Computational Social Science*, 07-11 September 2015, pp. 733-742.

[20] Lindsay T. Hoyt, PhD, Michelle G. Craske, PhD, Susan Mineka, PhD, and Emma K. Adam, PhD, "Positive and Negative Affect and Arousal : Cross-Sectional and Longitudinal Associations with Adolescent Cortisol Diurnal Rhythms", *Psychosomatic Medicine*, Vol. 77 No. 4, May 2015, pp. 392-401.

[21] Van Dongen B.F., Adriansyah A. "Process Mining: Fuzzy Clustering and Performance Visualization", in *Business Process Management Workshops,* Vol. 43, 2009, pp. 158-169.

[22] Arne Vikan, *A Fast Road to the Study of Emotions: Individual Emotion.* Springer, Cham, 2017, pp. 19-24.

[23] Kroenke, K., Spitzer, R. L., & Wiliams, J. B, "The PHQ-9: validity of a brief depression severity measure", *Journal of General Internal Medicine* , Vol. 16 No. 9, 2001.

[24] Kroenke, K., & Spitzer, R. L.,"The PHQ-9: A new depression diagnostic and severity measure", *Psychiatric Annals*, Vol. 32 No. 9, 2002, pp. 509-515.

[25] Lisa FeldmanBarrett, ElizaBliss Moreau, "Affect as a Psychological Primitive", *Advances in Experimental Social Psychology*, Vol. 41, 2009, pp. 167-218.

[26] Russell, J. A., Weiss, A., & Mendelsohn, G. A., "Affect Grid: A Single-Item Scale of Pleasure and Arousal", *Journal of Personality and Social Psychology*, Vol. 57 No. 3, 1989, pp. 493-502.

[27] Porouhan, P., & Premchaiswadi, W., "Pattern mining and process modelling of collaborative interaction data in an online multi-tabletop learning environment", *International Journal of Knowledge Engineering and Data Mining*, Vol. 4 No. 2, 2017, pp. 1-31.

[28] Porouhan, P., & Premchaiswadi, W., "Process Mining and Learners' Behavior Analytics in a Collaborative and Web-Based Multi-Tabletop Environment", *International Journal of Online Pedagogy and Course Design*, Vol. 7 No. 3, 2017, pp. 1-25.

[29] Premchaiswadi, W., & Porouhan, P., "Process modeling and bottleneck mining in online peer-review systems", *SpringerPlus (Computer Science)*, Vol. 4 No. 441, 2015, pp. 1-18.

[30] Premchaiswadi, W., & Porouhan, P., "Process modeling and decision mining in a collaborative distance learning environment", *Decision Analytics,* Vol. 2 No. 6, 2015, pp. 1-34.

[31] Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D.,"On the role of positive and negative affectivity in job performance: A meta-analytic investigation". *Journal of Applied Psychology*, Vol. 94 No. 1, 2009, pp. 162-176.