Chapter 05

Identifying User Interactions based parameters that affect emotions in e-Learning environment

5.1. Introduction

The objective of this chapter is to present the experimental design of the survey. In this project we intend to develop a model to determine the parameters that explain human emotions, based on human computer interactions that affect emotions, using multiple regression analysis.

The module that explains this is shown in Figure 6.1, The Top Level Architectural Design Diagram, under Emotions analysis module.

5.2 Model Equation

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + b_5 X_{5i} + b_6 X_{6i} + b_7 X_{7i} + b_8 X_{8i} + b_9 X_{9i} + e_i$$

Where

$Y_i =$ i th Observation of quiz marks obtained by student for a subject under study);
i=1 to N

$b_0 =$ Intercept and b_1 to b_7 **:** Slope coefficients for X_1 to X_7

e_i **=** Error term for the ith observation

X_1 = Deviation from expected time spent on a content page

X_2 = Deviation from expected time spent using application programs defined for the subject

X_3 = Relevancy of Applications used

X_4 = Time spent on Relevant Applications used

X_5 = Content feedback Index

X_6 = Emotions feedback index

X_7 = Deviation from total subject duration

X_8 = Keyword usage index

X_9 = Profile rating

5.3 Parameters

5.3.1 Deviation from expected time spent on a content page - X_1

For each subject, subject content is provided as pages. The lecturer teaching the subject assigns expected time to be spent on a page by an average student. The parameter under observation for a student is obtained by the difference between the expected time and the actual time.

$$X_1 = E(X_1) - X_{1\,\mathrm{Actual}}$$

Where,

$E(X_1)$ = Expected time spent on a content page.

$X_{1\,\mathrm{Actual}}$ =Actual time spent on a content page by the student

5.3.2 Deviation from expected time spent using application program groups defined for the subject X_2

For each subject, the lecturer defines application programs to be used by the student in order to enhance the skills. These application programs are grouped in to classes of applications, such as text editors, spread sheet applications or web browsers. He assigns expected time to be spent on a particular application program group by an average student. Table 5.1 shows sample data for such instance.

Subject	**Application Group**	**Application**	**Expected Time $E(X_2)$**
Subject01	AppGroup01	App01	T01
		App02	
		App03	
	AppGroup02	App04	T02
		App05	
Subject02	AppGroup01	App01	T'01
		App02	

Table 5.1: Application group usage time

The parameter under observation for a student is obtained by the difference between the expected time and the actual time for an application group.

$$X_2 = \sum_{i=1}^{n} E(X_2) - X_{2\,\mathrm{Actual}}$$

Where,

$E(X_2)$ = Expected time spent on an application program group.

$X_{2\,\mathrm{Actual}}$ = Actual time spent on an application program group by the student

5.3.3 Relevancy of Number of Applications used X_3

For each subject, the lecturer defines as application programs to be used, which are relevant to the subject as per the Table 5.1. These application programs are again grouped in to classes of applications. In the event a student using an application that is not in this list, it is considered as an irrelevant application. The next parameter under observation is the relevancy of application given by the following equation.

$$X_3 = \frac{\text{Number of relevant applications}}{\text{Total number of applications used}}$$

5.3.4 Time spent on Relevant Applications used X_4

For each subject, the lecturer defines as application programs to be used, and the expected time to be spent on an application. These application programs are grouped in to classes of applications. In the event a student using an application that is not in this list, it is considered as an irrelevant application. The next parameter under observation is the relevancy of application given by the following equation.

$$X_4 = \frac{\sum_{i=1}^{n} T_{iRelevant}}{\sum_{i=1}^{m} T_i}$$

Where $T_{iRelevant}$ = Time spent on i^{th} relevant applications

n= number of relevant applications used

T_i=Time spent on i^{th} application

m= number of applications used

5.3.5 Content feedback Index X_5

Popup feedback requests are used to obtain feedback on student's understanding on the subject that he had learned up to that particular point, after providing a set of content pages to the learner. Theses feedback questions, answers, options presented and the ratings are provided by the lecturer. Options are rated according to the understanding of the subject. For example, there will be an option that is totally irrelevant to the subject and if the student selects it, he will be given a low rating and vice versa. This is further explained in examples given in Table 5.2 and Table 5.3 below.

Subject	Content Page	Feedback Options to select	Rating
Subject01	ContentPage07	F01 F02 F03 F04	1 2 3 4
	ContentPage15	F01 F02 F03 F04	1 2 3 4
	ContentPage25	F01 F02 F03 F04	1 2 3 4
…			

Table 5.2: Feedback rating setting for subject content page

Student	Subject	Content Page	Feedback Received	Rating Obtained (Rating)
Student01	Subject01	ContentPage07	F02	1
		ContentPage15	F03	3
		ContentPage25	F01	1

Table 5.3: Example of feedback rating obtained by a student

$$X_5 = \sum_{i=1}^{n} \text{Rating}$$

Where Rating is the feedback rating received by the student at each content page that requires feedback information

5.3.6 Emotions feedback index X_6

Student's direct feedback on emotions while studying is obtained at different phases of study, in a similar manner discussed above. Data is gathered in the same way, assigning ratings to emotions. An example to derive at the parameter is given in the Table 5.4.

Student	Subject	Content Page	Emotion Feedback Received	Rating Obtained (Emotion Rating)
Student01	Subject01	ContentPage07	F02	1
		ContentPage15	F03	3
		ContentPage25	F01	1

Table 5.4: Example of feedback rating obtained by a student

$$X_6 = \sum_{i=1}^{n} \text{EmotionRating}$$

5.3.7 Deviation from total subject duration X_7

Subject duration for all subjects is defined when courses are prepared. The parameter under observation for a student is obtained by the difference between the expected time and the actual time.

$$X_7 = E(X_7) - X_{7\,\text{Actual}}$$

Where,

$E(X_7)$ = Expected time spent on the subject.

$X_{7\,\text{Actual}}$ = Actual time spent on the subject by the student

5.3.8 Keyword usage index X_8

A set of keywords identified by the lecturer that are relevant to the subject are stored in the system.

As the first step of using this information as parameters, we intend to find the percentage of keywords typed in by the student. Prepositions, conjunctions and similar words such as “in, on, and, the, a “typed will be excluded from this count.

$$X_8 = \frac{\text{Total number of keywords used}}{\text{Total number of words used}}$$

5.3.9 Profile rating X_9

Students are rated by the lecturer based on their expertise and exposure to the subject. This is also used as a parameter in this study.

X_9 = Student rating

The model is developed using the above available information for the two years 2005 and 2006.

Estimated Multiple Regression (MR) formula was obtained by

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} + \hat{b}_3 X_{3i} + \hat{b}_4 X_{4i} + \hat{b}_5 X_{5i} + \hat{b}_6 X_{6i} + \hat{b}_7 X_{7i} + \hat{b}_8 X_{8i} + \hat{b}_9 X_{9i}$$

5.4 Results

Analyzing the interactions related parameters using Microsoft Excel produces following results given in Table 5.5, Table 5.6 and Table 5.7.

Regression Statistics	
Multiple R	
R Square	
Adjusted R Square	
Standard Error	
Observations	*N*

Table 5.5: Regression Results

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	n				
Residual	N-n-1				
Total					

Table 5.6: ANOVA

	Coefficients	*Standard Error*	*t Statistic*
b_0			
X_1			
X_2			
X_3			
X_4			
X_5			
X_6			
X_7			
X_8			
X_9			

Table 5.7: t-statistic table

Hence, the regression model is given as follows

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} + \hat{b}_3 X_{3i} + \hat{b}_4 X_{4i} + \hat{b}_5 X_{5i} + \hat{b}_6 X_{6i} + \hat{b}_7 X_{7i} + \hat{b}_8 X_{8i} + \hat{b}_9 X_{9i}$$

5.5 Statistical Testing

5.5.1 Overall Effectiveness of the Model

Overall effectiveness of X_i in explaining Y_i is tested using the following

5.5.1.1 Multiple Coefficient of Determination R^2 & Adjusted Multiple Coefficient of Determination R_a^2

R^2 = rr indicates that the model, as a whole, explains rr% of the variation in the dependant variable. Also $R^2 \approx R_a^2$ indicates that R^2 is not dependant on the number of independent variables

5.5.1.2 F - Statistic

F – Statistic is used to test how well a set of independent variables collectively explain the variation in the dependant variable. That is, the F statistic is used to test whether at least one X_i (i = 1 to 9) explains a significant portion of the variation of Y_i.

Null hypothesis,
H_0: $b_1 = b_2 = b_3 = b_4 = b_5 = b_6 = b_7 = b_8 = b_9 = 0$ versus alternative hypothesis
H_a: at least one $b_i \neq 0$.

The decision rule to reject H_0 is $F > F_C$, where F = calculated F value, and F_C = Critical F value from F table. The outcome of the test is as follows.

F = xx
F_C = yy at 95% significance level for numerator degrees of freedom = n and denominator degrees of freedom = N-n,

If $F > F_C$ the null hypothesis is rejected at 5% level of significance, i.e., at least one of the seven X_i (i = 1 to 9) is significantly different than 0.

5.5.2 Significance of the individual independent variables

The contribution of individual variables in explaining the dependent variable is indicated by the significance of the slope coefficient. This is tested by the t-statistic.

H_0: $b_i = 0$ versus alternative hypothesis
H_a: $b_i \neq 0$,
for i= 1 to 9

The decision rule; to reject H_0 if t_{b_i} > $+t_c$, or t_{b_i} < $-t_c$ where t_{b_i} = calculated t value, and t_c = critical t value from t table.

The outcome of the test is given in Table 5.8.

$t_c = \pm 1.96$ for 5% significant level at N-n degrees of freedom (two tailed test)

Example

	t Stat	*Result*	*Interpretation*
b_0		> 1.96	Reject H_0
b_1			Cannot reject H_0
b_2		< -1.96	Reject H_0
b_3		> 1.96	Reject H_0
b_4		> 1.96	Reject H_0
b_5		< -1.96	Reject H_0
b_6			Cannot reject H_0
b_7			Cannot reject H_0
b_8			Cannot reject H_0
b_9			Cannot reject H_0

Table 5.8: Significance of individual variables

Hence, it is evident that the following variables have slope coefficients that are significantly different from 0 at the 5% level of significance, indicating that they contribute to explaining the Quiz Marks.

X_2 = Deviation from expected time spent using application programs defined for the subject

X_3 = Relevancy of Applications used

X_4 = Time spent on Relevant Applications used

X_5 = Content feedback Index

In addition, the intercept term is also statistically significant.

It is also noted that following variables do not contribute to explain Quiz marks at 5% level of significance.

X_1 = Deviation from expected time spent on a content page
X_6 = Emotions feedback index
X_7 = Deviation from total subject duration
X_8 = Keyword usage index
X_9 = Profile rating

5.5.3 Detecting Serial Correlation (Autocorrelation)

This is the situation in which the residual terms are correlated with one another. Durbin - Watson statistic is used to detect this. Durbin – Watson Statistic (DW) is given by,

$$\boxed{DW = \frac{[\sum_{i=2}^{N}(\hat{e}_i - \hat{e}_{i-1})^2]}{\sum_{i=1}^{N}(\hat{e}_i)^2}}$$

Where $\hat{e}_t$ =residual for student i.

Hypothesis test is carried out taking
H_o: regression has no serial correlation versus alternative hypothesis
H_a: regression has serial correlation

Using Durbin – Watson Statistic (DW), the calculated value for the test statistic is dw-calc. An example is shown in Figure 5.1.

Reject H_o Positive serial Correlation	Inconclusive	do not reject H_o	Inconclusive	Reject H_o Negative serial Correlation

0 $\quad d_l \quad d_u \quad 4-d_u \quad 4-d_l \quad$ 4

Figure 5.1: Durbin – Watson statistic hypothesis

If 0< DW =dw-calc < d_l = 1.57, the H_o is rejected.

5.5.4 Testing for Heteroskedasticity

One of the assumptions of the Multiple Regression is that the residual terms are constant across observations. Heteroskedasticity occur when this is not so. This leads to standard errors of coefficients to differ for different values of the independent variables and influence the conclusions of t- tests on them.

Breusch – Pagan (BP) test statistic is used to detect heteroskedasticity.

H_o: regression has no problem in heteroskedasticity versus alternative hypothesis
H_a: regression has a problem in heteroskedasticity

BP = n*R^2 with k degrees of freedom, where
n = no of observations = N
R^2 = Multiple coefficient of determination = rr and
k = no of independent variables = 9

Estimated BP = bp-calc

One tailed critical value for a chi-squared distribution with 9 degrees of freedom at 5% significance level = chi-squared-calc

If BP > critical value(chi-squared-calc), H_o is rejected, i.e. there is a problem with heteroskedasticity.

5.5.5 Detecting Multicollinearity

Multicolliniarity rejects the condition under which a high correlation exists among two or more independent variables in the multiple regression equation. Analysis Results are shown in Table 5.9.

		Coefficients	P-value	Result
b0	*Intercept*		*0%*	
b1	*X Variable 1*		*40%*	*> 10%*
b2	*X Variable 2*		*0%*	
b3	*X Variable 3*		*0%*	
b4	*X Variable 4*		*4%*	
b5	*X Variable 5*		*0%*	
b6	*X Variable 6*		*56%*	*> 10%*
b7	*X Variable 7*		*99%*	*> 10%*
b8	*X Variable 8*			
b9	*X Variable 9*			
F - Test		*FF*	*Significant*	
R^2		*rr%*	*High*	

Table 5.9: Multicolleniarity table

- t test indicates 4 out of 9 variables are significantly different from zero.
- F statistic is statistically significant
- R^2 is relatively high

The above suggests that we can not totally exclude the possibility of the existence of Multicolliniarity.

5.6 Conclusion

1. If R^2 and R_a^2, and also by F Statistic, the model has a relatively high explanatory power.
2. However, if t test indicates, 5 out of 9 independent variables identified are not statistically significant (in the example shown here).
3. The model is threatened with heteroskedasticity and serial correlation (in the example given above). Presence of multicolliniarity also can not be excluded totally.

Therefore it is recommended that the model specification be changed and tested again, i.e. the statistically insignificant variables must be removed and tested again till a suitable result is obtained.