

**PENILAIAN ESEI BERBANTUKAN KOMPUTER MENGGUNAKAN  
TEKNIK BAYESIAN DAN PENGUNDURAN LINEAR BERGANDA**

**MOHD AZWAN BIN MOHAMAD@HAMZA**

**UNIVERSITI TEKNOLOGI MALAYSIA**

*Khas buat ibu, abah, isteri, puteri dan bakal puteriku yang dikasih...*

## PENGHARGAAN

Dengan Nama Allah Yang Pemurah Lagi Maha Pengasihani.

Assalamualaikum W.B.T...

Segala puji-pujian bagi Allah S.W.T, Tuhan semesta alam. Salawat dan salam ke atas junjungan besar Nabi Muhammad S.A.W, keluarga dan para sahabat baginda serta kaum Muslimin dan Muslimat.

Bersyukur saya ke hadrat Allah S.W.T kerana di atas limpah kurnia-Nya serta keizinan-Nya, dapatlah jua saya menyiapkan Laporan Projek Sarjana Muda ini.

Di kesempatan ini juga ingin saya merakamkan jutaan terima kasih dan penghargaan ikhlas buat PM Abdul Manan b Ahmad, selaku penyelia bagi projek ini, atas bimbingan dan dorongan yang diberikan sepanjang tempoh penyelidikan projek ini.

Kerjasama yang baik daripada pihak Sekolah Kebangsaan Galing dan Sekolah Kebangsaan Semambu, Kuantan serta pihak Dewan Bahasa dan Pustaka turut dihargai sehingga saya dapat memperoleh maklumat yang diingini yang mana ianya menyumbang kepada sebahagian besar kejayaan pembangunan projek ini.

Penghargaan juga turut ditujukan kepada semua yang terlibat samada secara langsung atau tidak langsung dalam membantu menjayakan projek penyelidikan ini.

## ABSTRAK

Perbezaan markah antara dua penilai, peruntukan masa yang panjang dan kos pemarkahan yang tinggi menjadi punca yang menyebabkan Penilaian Esei Berbantuan Komputer (CbAS) dikaji. Kunci utama ialah penilaian CbAS mestilah hampir setara dengan penilaian manusia. Berdasarkan skema penilaian esei UPSR, terdapat tiga komponen utama penilaian iaitu bahasa, elemen hujahan (isi kandungan) dan gaya olahan. Didapati penggunaan Logik Fuzzy dalam menentukan dan mengelaskan elemen hujahan dan Algoritma Pengunduran Linear *Stepwise* (SLR) dalam membuat peramalan terhampir bagi gaya olahan masih terdapat beberapa kelemahan. Logik Fuzzy tidak mengukur bentuk ciri bahasa dan memerlukan saiz data latihan yang besar. Manakala Algoritma SLR menghasilkan peramalan gaya olahan menggunakan ciri yang kurang piawai di samping saiz set ciri tidak ditakrifkan dengan jelas dan tiada jaminan ciri yang dipilih adalah signifikan untuk menyumbang kepada peramalan gaya olahan terhampir. Kajian ini memberi penekanan ke atas peramalan elemen hujahan dan gaya olahan yang lebih optimum yang mendorong kepada pembangunan CbAS berdasarkan empat fasa metodologi penyelidikan. (1) Fasa pra-pemprosesan dan pengekstrakan data di mana esei dipecahkan kepada token (perkataan) dan menggunakan Algoritma Pembetulan Kata membetulkan ejaan yang salah. (2) Fasa proses latihan penentuan dan pengelasan elemen hujahan menggunakan Teknik Model *Multivariate Bernoulli* (MMB) yang mengambilkira ciri yang wujud dan tidak wujud seterusnya mengukur bentuk ciri bahasa yang mempengaruhi kualiti esei tersebut. Teknik MMB juga hanya memerlukan saiz korpus data yang lebih kecil. (3) Fasa proses peramalan gaya olahan dengan menggunakan Algoritma Pengunduran Linear Berganda (MLR). Algoritma MLR menggunakan enam ciri yang telah ditetapkan (berdasarkan kajian terdahulu) supaya peramalan yang dibuat lebih piawai dan set ciri tersebut adalah lebih signifikan. (4) Fasa pengujian kesetaraan pencapaian daripada gabungan MMB, MLR dan data bahagian bahasa (dari penilaian manusia) dan dibandingkan dengan penilaian manusia untuk lima kitaran *cross-validation*. Hasil menunjukkan pencapaian adalah konsisten dengan peratus kesetaraan iaitu 95.2%. Kesimpulannya, eksperimen menunjukkan dengan menggunakan kedua-dua teknik (MMB dan MLR), peramalan atau penilaian esei yang lebih baik telah dicapai berbanding dengan sistem yang menggunakan Logik Fuzzy dan Algoritma SLR.

## ABSTRACT

Disagreement of grade given by two human judges, time consuming and high evaluation cost became a reason of research on Computer-based Assessment System (CbAS) been studied. The main key is CbAS assessment must be closest to human assessment. Based on UPSR Essay Assessment Schema, there are three main assessment components consists of language, discourse element and style. Recently, Fuzzy Logic is used to determine and classify the discourse element while Stepwise Linear Regression Algorithm (SLR) is used to make closest prediction for style of writing. Both of them have its weakness. Fuzzy Logic did not measure the form of linguistic features and required a huge size of training data. SLR Algorithm derive prediction of writing style using un-standardize feature set and size of features set not clearly defined and no warranty of significance in contribute to get closest grade prediction. This study emphasized on optimization of prediction on discourse elements and writing style that leading to the development of CbAS through four phases of research methodology. (1) Pre-processing and data extraction phase where essay will be parsed into word (token) and implemented Word Correction Algorithm to re-correct the misspell word. (2) Training process of determination and classification of discourse elements using Multivariate Bernoulli Model (MMB) Technique. It considers both presence and absence features thus it measured the form of linguistic features that reflected essay quality. MMB Technique only required a small size of training data. (3) Prediction process of writing style using Multiple Linear Regression (MLR) Algorithm. MLR Algorithm applied six fixed features (based on previous research) to ensure the prediction is more standardize and feature set is more significant. (4) Test the performance agreement derived from the combination of MMB, MLR and data of language component (taken from human assessment) and compared it to human assessment for five cycles of cross-validation. The outcome shows performance is consistent with 95.2% agreement. Thus, the experiment has shown by utilizing both techniques (MMB and MLR), better prediction or essay assessment has been achieved compared to the one's implemented using Fuzzy Logic and SLR Algorithm.

## ISI KANDUNGAN

| <b>BAB</b> | <b>PERKARA</b>                     | <b>MUKA SURAT</b> |
|------------|------------------------------------|-------------------|
|            | <b>JUDUL</b>                       | <b>i</b>          |
|            | <b>PENGAKUAN</b>                   | <b>ii</b>         |
|            | <b>DEDIKASI</b>                    | <b>iii</b>        |
|            | <b>PENGHARGAAN</b>                 | <b>iv</b>         |
|            | <b>ABSTRAK</b>                     | <b>v</b>          |
|            | <b>ABSTRACT</b>                    | <b>vi</b>         |
|            | <b>ISI KANDUNGAN</b>               | <b>vii</b>        |
|            | <b>SENARAI JADUAL</b>              | <b>xiii</b>       |
|            | <b>SENARAI RAJAH</b>               | <b>xv</b>         |
|            | <b>SENARAI RUMUS DAN ALGORITMA</b> | <b>xviii</b>      |
|            | <b>SENARAI SINGKATAN</b>           | <b>xx</b>         |
|            | <b>SENARAI ISTILAH</b>             | <b>xxii</b>       |
|            | <b>SENARAI LAMPIRAN</b>            | <b>xxiv</b>       |
| <b>1</b>   | <b>Pengenalan</b>                  | <b>1</b>          |
|            | 1.1 Pendahuluan                    | 1                 |
|            | 1.2 Latar Belakang Masalah         | 3                 |
|            | 1.3 Pernyataan Masalah             | 8                 |
|            | 1.4 Matlamat                       | 8                 |
|            | 1.5 Objektif                       | 9                 |

|          |  |           |
|----------|--|-----------|
| 1.6      | Skop   | 9         |
| 1.7      | Kepentingan Penyelidikan                             | 10        |
| 1.8      | Sumbangan Ilmiah                                     | 10        |
| 1.9      | Struktur Tesis                                       | 11        |
| 1.10     | Ringkasan  | 12        |
| <b>2</b> | <b>KAJIAN LITERATUR</b>                              | <b>14</b> |
| 2.1      | Pendahuluan  | 14        |
| 2.2      | Latar Belakang Penilaian Esei Berbantuan Komputer    | 16        |
| 2.2.1    | Kajian Awal  | 16        |
| 2.2.2    | Kajian Semasa  | 18        |
| 2.2.2.1  | Project Essay Grader (PEG)                           | 20        |
| 2.2.2.2  | Intelligent Essay Assessor (IEA)                     | 22        |
| 2.2.2.3  | Educational Testing Service (ETS I)                  | 24        |
| 2.2.2.4  | Electronic Essay Rater (E-rater)                     | 27        |
| 2.2.2.5  | Conceptual Rater (C-rater)                           | 29        |
| 2.2.2.6  | Bayesian Essay Test Scoring sYstem<br>(BETSY)        | 30        |
| 2.2.2.7  | Intelligent Essay Marking Systems<br>(IEMS)          | 33        |
| 2.2.2.8  | Automark   | 33        |
| 2.2.2.9  | Schema Extract Analyse and Report<br>(SEAR)          | 35        |
| 2.2.2.10 | Paperless School free-text Marking<br>Engine (PS-ME) | 36        |
| 2.2.3    | Isu/Analisa Kajian                                   | 38        |
| 2.2.4    | Cadangan Teknik Penilaian                            | 45        |
| 2.2.5    | Set Ciri Peramalan                                   | 46        |
| 2.3      | Pemarkahan Esei                                      | 49        |
| 2.3.1    | Peraturan Memberi Markah (PMM)                       | 49        |
| 2.3.1.1  | Teknik Global  | 50        |
| 2.3.1.2  | Teknik Analisis                                      | 51        |
| 2.3.2    | Kaedah Menilai Esei                                  | 52        |

|          |  |           |
|----------|--|-----------|
| 2.3.3    | Penyelarasan Markah                                  | 52        |
| 2.3.4    | Pemarkahan Esei UPSR                                 | 53        |
| 2.3.5    | Isu-isu Utama Dalam Penilaian Esei                   | 55        |
| 2.3.5.1  | Penandaan Kesalahan                                  | 56        |
| 2.3.5.2  | Keadaan Kritikal                                     | 56        |
| 2.4      | Ringkasan  | 56        |
| <b>3</b> | <b>METODOLOGI KAJIAN</b>                             | <b>58</b> |
| 3.1      | Pendahuluan  | 58        |
| 3.2      | Metodologi Penyelidikan                              | 59        |
| 3.3      | Teknik Penilaian                                     | 60        |
| 3.4      | Prosidur <i>Five-Fold Cross-Validation</i>           | 61        |
| 3.5      | Set 12 Ciri Peramalan                                | 63        |
| 3.5.1    | Set Ciri Optimum                                     | 65        |
| 3.5.1.1  | Purata Bilangan Perkataan dalam Elemen Hujahan (AEL) | 65        |
| 3.5.1.2  | Jenis/Token  | 65        |
| 3.5.1.3  | Purata Panjang Perkataan (AWL)                       | 66        |
| 3.5.1.4  | Kandungan Esei (EC)                                  | 66        |
| 3.5.1.5  | Kandungan Argumen (AC)                               | 66        |
| 3.5.1.6  | Panjang Esei (EL)                                    | 67        |
| 3.5.2    | Set Ciri Tetap                                       | 67        |
| 3.5.2.1  | Kesalahan Bahasa                                     | 67        |
| 3.5.2.2  | Elemen Hujahan                                       | 68        |
| 3.6      | Pra-Pemprosesan Data dan Pengekstrakan Data          | 69        |
| 3.7      | Latihan Peramalan Gaya Olahan (Ciri Optimum)         | 72        |
| 3.7.1    | Pemilihan Ciri                                       | 73        |
| 3.7.2    | Algoritma MLR  | 73        |
| 3.7.2.1  | Matrik Pelengkap                                     | 74        |
| 3.7.2.2  | Pendaraban Silang Matrik                             | 75        |
| 3.7.2.3  | Matrik Songsang                                      | 75        |
| 3.7.3    | Pekali Pemberat Pengunduran                          | 75        |
| 3.8      | Latihan Pengkelasan Elemen Hujahan (Ciri Tetap)      | 76        |



|          |  |           |
|----------|--|-----------|
| 3.8.1    | Model <i>Multinomial</i>   | 76        |
| 3.8.2    | Model <i>Multivariate Bernoulli</i>  | 77        |
| 3.8.3    | Pemilihan Data   | 78        |
| 3.8.3.1  | Kriteria Kedudukan, $A_1$  | 79        |
| 3.8.3.2  | Kriteria Leksikal, $A_2$   | 79        |
| 3.8.3.3  | Kriteria Teori Struktur Retorik, $A_3$   | 80        |
| 3.8.4    | Penapisan Data   | 82        |
| 3.8.5    | Pekali Pemberat Peratusan  | 83        |
| 3.9      | Pengujian Penilaian  | 83        |
| 3.9.1    | Pengujian Ciri Tetap   | 83        |
| 3.9.1.1  | Ciri Bahasa  | 84        |
| 3.9.1.2  | Teknik MMB   | 84        |
| 3.9.1.3  | Pekali Pemberat Peratusan  | 85        |
| 3.9.2    | Pengujian Ciri Optimum   | 86        |
| 3.9.2.1  | Algoritma MLR  | 86        |
| 3.9.2.2  | Pekali Pemberat Pengunduran  | 87        |
| 3.9.2.3  | Penilaian Gred Akhir   | 87        |
| 3.10     | Ringkasan  | 88        |
| <b>4</b> | <b>TEKNIK PENILAIAN MENGGUNAKAN BAYESIAN<br/>DAN PENGUNDURAN LINEAR BERGANDA</b> | <b>89</b> |
| 4.1      | Pendahuluan  | 89        |
| 4.2      | Rekabentuk Pangkalan Pengetahuan (KB)  | 89        |
| 4.3      | Teknik Penentuan dan Pengelasan Elemen Hujahan                                   | 90        |
| 4.3.1    | Model <i>Multivariate Bernoulli</i> (MMB)  | 90        |
| 4.3.1.1  | Kebarangkalian <i>Prior</i>  | 92        |
| 4.3.1.2  | Kebarangkalian <i>Conditional</i>  | 92        |
| 4.3.2    | Logik Fuzzy  | 94        |
| 4.3.2.1  | Penapisan Token  | 94        |
| 4.3.2.2  | Set Fuzzy  | 94        |
| 4.3.2.3  | Darjah Keahlian  | 95        |
| 4.3.2.4  | Operasi Set Fuzzy  | 97        |
| 4.3.2.5  | <i>Cartesian Product</i>   | 98        |

|          |  |            |
|----------|--|------------|
| 4.3.2.6  | Hubungan Fuzzy                                   | 98         |
| 4.3.2.7  | Operasi <i>Compositional</i>                     | 99         |
| 4.4      | Teknik Peramalan Gaya Olahan                     | 99         |
| 4.4.1    | Model Pengunduran Linear Berganda                | 99         |
| 4.4.1.1  | Peramalan Parameter                              | 100        |
| 4.4.2    | Model Pengunduran Linear <i>Stepwise</i>         | 103        |
| 4.4.2.1  | Algoritma Pemilihan Pembolehubah                 | 103        |
| 4.5      | Prosidur Penilaian                               | 106        |
| 4.5.1    | Penjelmaan Linear Ciri Tetap                     | 106        |
| 4.5.2    | Penentuan Nilai Ciri Optimum                     | 106        |
| 4.5.3    | Penentuan Pekali Pemberat Ciri Optimum           | 108        |
| 4.5.4    | Peratusan Pemberat Ciri Tetap                    | 109        |
| 4.5.5    | Penentuan Pekali Pemberat Ciri Tetap             | 109        |
| 4.5.6    | Penilaian Gred Akhir                             | 110        |
| 4.6      | Ringkasan  | 110        |
| <b>5</b> | <b>HASIL PENGUJIAN DAN PERBINCANGAN</b>          | <b>112</b> |
| 5.1      | Pendahuluan                                      | 112        |
| 5.2      | Hasil Pengujian dan Perbincangan                 | 112        |
| 5.3      | Kaedah Pengukuran                                | 113        |
| 5.4      | Hasil Pra-Pemprosesan Data                       | 114        |
| 5.4.1    | Rumusan Pra-Pemprosesan Data                     | 116        |
| 5.5      | Hasil Peramalan Gaya Olahan                      | 117        |
| 5.5.1    | Rumusan Peramalan Gaya Olahan                    | 125        |
| 5.6      | Hasil Penentuan dan Pengkelasan Elemen Hujahan   | 126        |
| 5.6.1    | Rumusan Penentuan dan Pengkelasan Elemen Hujahan | 134        |
| 5.7      | Hasil Teknik Penilaian                           | 136        |
| 5.8      | Rumusan Keseluruhan                              | 138        |
| 5.9      | Ringkasan  | 138        |

|          |  |            |
|----------|--|------------|
| <b>6</b> | <b>KESIMPULAN</b>  | <b>140</b> |
| 6.1      | Pendahuluan  | 140        |
| 6.2      | Kesimpulan   | 140        |
| 6.3      | Cadangan Kajian Lanjutan                                   | 142        |
| 6.3.1    | Kaedah Gabungan Kata dan Penapisan Ciri                    | 143        |
| 6.3.2    | Penulisan Di Luar Topik                                    | 143        |
| 6.4      | Ringkasan  | 144        |
|          | <br>   |            |
|          | RUJUKAN  | 145        |
|          | LAMPIRAN A : Contoh Data Pengujian                         | 158        |
|          | LAMPIRAN B : Contoh Data Latihan                           | 171        |
|          | LAMPIRAN C : Skema Penilaian Esei UPSR                     | 184        |
|          | LAMPIRAN D : Skala Panduan Markah Gaya Olahan              | 186        |
|          | LAMPIRAN E : Senarai Keseluruhan Ciri Untuk Penilaian Esei | 188        |

## SENARAI JADUAL

| NO. JADUAL | TAJUK  | MUKA SURAT |
|------------|--|------------|
| 2.1        | Pengkelasan Sistem Penilaian Esei Berautomasi..                                  | 39         |
| 2.2        | Perbandingan pencapaian sistem penilaian.  | 41         |
| 2.3        | Set 12 ciri peramalan untuk teknik penilaian.                                    | 48         |
| 2.4        | Pembahagian markah berdasarkan Skema Penilaian Esei UPSR.                        | 55         |
| 3.1        | Kitaran prosidur <i>five-fold cross validation</i> .                             | 62         |
| 3.2        | Set 12 ciri signifikan untuk peramalan penilaian.                                | 64         |
| 3.3        | Gred markah UPSR.  | 87         |
| 4.1        | Darjah keahlian bagi Set Fuzzy A dan Set Fuzzy B.                                | 95         |
| 4.2        | Data untuk Pengunduran Linear Berganda.  | 101        |
| 5.1        | Pecahan bilangan 200 sampel esei pelajar mengikut gred.                          | 113        |
| 5.2        | Perbandingan perkataan diperbetulkan berdasarkan gred esei skala lima-mata aras. | 115        |
| 5.3        | Hasil ujian peramalan gaya olahan pada kitaran pertama.                          | 118        |
| 5.4        | Hasil ujian peramalan gaya olahan pada kitaran kedua.                            | 120        |

|      |  |     |
|------|--|-----|
| 5.5  | Hasil ujian peramalan gaya olahan pada kitaran ketiga.               | 121 |
| 5.6  | Hasil ujian peramalan gaya olahan pada kitaran keempat.              | 122 |
| 5.7  | Hasil ujian peramalan gaya olahan pada kitaran kelima.               | 124 |
| 5.8  | Hasil penentuan dan pengkelasan elemen hujahan pada kitaran pertama. | 127 |
| 5.9  | Hasil penentuan dan pengkelasan elemen hujahan pada kitaran kedua.   | 128 |
| 5.10 | Hasil penentuan dan pengkelasan elemen hujahan pada kitaran ketiga.  | 130 |
| 5.11 | Hasil penentuan dan pengkelasan elemen hujahan pada kitaran keempat. | 131 |
| 5.12 | Hasil penentuan dan pengkelasan elemen hujahan pada kitaran kelima.  | 133 |
| 5.13 | Hasil perbandingan penilaian manusia dan teknik penilaian.           | 136 |

## SENARAI RAJAH

| NO. RAJAH | TAJUK   | MUKA SURAT |
|-----------|---|------------|
| 2.1       | Garisma perkembangan kajian dalam bidang penilaian penulisan.               | 17         |
| 2.2       | Versi demo Sistem PEG yang dipaparkan di laman web.                         | 21         |
| 2.3       | Contoh maklumbalas yang diberikan oleh Sistem IEA.                          | 23         |
| 2.4       | Pepohon sintaksis X-bar.  | 26         |
| 2.5       | Antaramuka ramah-pengguna Sistem E-rater versi berpandukan-web (Criterion). | 28         |
| 2.6       | Pembangunan konseptual sistem penilaian.                                    | 40         |
| 3.1       | Rekabentuk metodologi penyelidikan.   | 59         |
| 3.2       | Contoh esei pelajar.  | 62         |
| 3.3       | Algoritma Pembetulan Kata.  | 71         |
| 3.4       | Proses latihan peramalan gaya olahan.                                       | 72         |
| 3.5       | Algoritma Pengunduran Linear Berganda.                                      | 74         |
| 3.6       | Proses latihan pengelasan elemen hujahan.                                   | 76         |
| 3.7       | Contoh data latihan menggunakan Kriteria Kedudukan.                         | 79         |
| 3.8       | Contoh data latihan menggunakan Kriteria Leksikal.                          | 80         |
| 3.9       | Pepohon RST.  | 81         |
| 3.10      | Contoh data latihan menggunakan Kriteria RST.                               | 82         |

|      |   |     |
|------|---|-----|
| 3.11 | Proses pengujian pengkelasan elemen hujahan.  | 84  |
| 3.12 | Proses pengujian peramalan gaya olahan.   | 86  |
| 4.1  | Pepohon Rangkaian MMB.  | 91  |
| 4.2  | Graf yang menunjukkan keahlian bagi Set Fuzzy A dan B.  | 97  |
| 5.1  | Perbandingan bilangan Ralat-Ejaan dan Ejaan-Diperbetulkan dengan kumpulan gred esei.                                  | 116 |
| 5.2  | Peratus Perkataan-Diperbetulkan berbanding kumpulan gred esei.  | 116 |
| 5.3  | Perbandingan hasil Algoritma MLR dan SLR pada kitaran pertama.  | 119 |
| 5.4  | Perbandingan hasil Algoritma MLR dan SLR pada kitaran kedua.  | 120 |
| 5.5  | Perbandingan hasil Algoritma MLR dan SLR pada kitaran ketiga.   | 122 |
| 5.6  | Perbandingan hasil Algoritma MLR dan SLR pada kitaran keempat.  | 123 |
| 5.7  | Perbandingan hasil Algoritma MLR dan SLR pada kitaran kelima.   | 124 |
| 5.8  | Rumusan perbandingan hasil Algoritma MLR dan SLR pada kelima-lima kitaran   | 126 |
| 5.9  | Peratus perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy pada kitaran pertama. | 128 |
| 5.10 | Peratus perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy pada kitaran kedua.   | 129 |
| 5.11 | Peratus perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy pada kitaran ketiga.  | 131 |

|      |  |     |
|------|--|-----|
| 5.12 | Peratus perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy pada kitaran keempat.  | 132 |
| 5.13 | Peratus perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy pada kitaran kelima.   | 134 |
| 5.14 | Rumusan perbandingan penentuan dan pengkelasan elemen hujahan antara Teknik MMB dan Logik Fuzzy bagi kelima-lima kitaran <i>five-fold cross-validation</i> . | 135 |
| 5.15 | Graf perbandingan penilaian gred esei antara penilaian manusia dan teknik penilaian bagi kelima-lima kitaran <i>five-fold cross-validation</i> .             | 137 |



## SENARAI RUMUS DAN ALGORITMA

| NO<br>ALGORITMA/<br>RUMUS | TAJUK   | MUKA<br>SURAT |
|---------------------------|---|---------------|
| 2.1                       | Rumus umum MMB.   | 31            |
| 2.2                       | Rumus umum MM.  | 32            |
| 4.1                       | Rumus MMB menggunakan <i>ln</i> kebarangkalian.                             | 91            |
| 4.2                       | Rumus kebarangkalian <i>prior</i> .   | 92            |
| 4.3                       | Rumus kebarangkalian <i>conditional</i> bagi kes pertama.                   | 93            |
| 4.4                       | Rumus kebarangkalian <i>conditional</i> bagi kes kedua.                     | 93            |
| 4.5                       | Rumus untuk mendapatkan kebarangkalian <i>conditional</i> bagi kes pertama. | 93            |
| 4.6                       | Rumus untuk mendapatkan kebarangkalian <i>conditional</i> bagi kes kedua.   | 93            |
| 4.7                       | Fungsi keahlian <i>Trapezoidal</i> bagi Set Fuzzy A.                        | 96            |
| 4.8                       | Fungsi keahlian <i>Trapezoidal</i> bagi Set Fuzzy B.                        | 96            |
| 4.9                       | Rumus Set Fuzzy A.  | 96            |
| 4.10                      | Operasi Set Fuzzy <i>Union</i> .  | 97            |
| 4.11                      | Operasi Set Fuzzy <i>Intersection</i> .                                     | 97            |
| 4.12                      | <i>Cartesian product</i> .  | 98            |
| 4.13                      | Subset Fuzzy kepada <i>Cartesian product</i> .                              | 98            |
| 4.14                      | Subset Fuzzy bagi $A \times B$ .  | 98            |
| 4.15                      | Matrik <i>R</i> .   | 99            |

|      |  |     |
|------|--|-----|
| 4.16 | Operasi <i>Compositional</i> .                       | 99  |
| 4.17 | Rumus Umum Pengunduran Linear Berganda.              | 100 |
| 4.18 | Bentuk skala bagi persamaan normal                   | 101 |
| 4.19 | Rumus normal kuasa dua-terkecil.                     | 101 |
| 4.20 | Rumus normal kuasa dua-terkecil dalam bentuk matrik. | 102 |
| 4.21 | Peramal kuasa dua terkecil                           | 102 |
| 4.22 | Peramal kuasa dua terkecil bagi $\beta$              | 102 |
| 4.23 | Peramal kuasa dua terkecil dalam bentuk matrik       | 102 |
| 4.24 | Vektor <i>residual</i>                               | 103 |
| 4.25 | Statistik $F^*$ untuk pengujian $F$ .                | 104 |
| 4.26 | Statistik ujian $F$ setara.                          | 104 |
| 4.27 | Penentuan pembolehubah $X$ .                         | 105 |
| 4.28 | Rumus penjelmaan linear.                             | 106 |
| 4.29 | Rumus kandungan esei.                                | 107 |
| 4.30 | Rumus pemberat perkataan.                            | 107 |
| 4.31 | Rumus kandungan argumen.                             | 108 |
| 4.32 | Rumus pemberat ciri tetap                            | 109 |
| 4.33 | Rumus gabungan penilaian                             | 110 |
| 4.34 | Ringkasan rumus gabungan penilaian                   | 110 |
| 5.1  | Rumus <i>precision</i> .                             | 114 |
| 5.2  | Rumus <i>recall</i> .                                | 114 |
| 5.3  | Rumus <i>f-measure</i> .                             | 114 |
| 5.4  | Rumus peramalan gaya olahan kitaran pertama.         | 117 |
| 5.5  | Rumus peramalan gaya olahan kitaran kedua.           | 117 |
| 5.6  | Rumus peramalan gaya olahan kitaran ketiga.          | 118 |
| 5.7  | Rumus peramalan gaya olahan kitaran keempat.         | 118 |
| 5.8  | Rumus peramalan gaya olahan kitaran kelima.          | 118 |

## SENARAI SINGKATAN

|         |   |  |
|---------|---|--|
| CbAS    | - | Penilaian Esei Berbantuan Komputer   |
| UPSR    | - | Ujian Penilaian Sekolah Rendah   |
| SLR     | - | Pengunduran Linear <i>Stepwise</i>   |
| MMB     | - | Model <i>Multivariate Bernoulli</i>  |
| MLR     | - | Pengunduran Linear Berganda  |
| AGREEM  | - | Kesetaraan penilaian antara penilaian manusia dan penilaian menggunakan teknik penilaian |
| AI      | - | Kepintaran Buatan  |
| GMAT    | - | <i>Graduate Management Admissions Test</i>   |
| NLP     | - | Pemprosesan Bahasa Tabii   |
| PEG     | - | <i>Project Essay Grader</i>  |
| IEA     | - | <i>Intelligent Essay Assessor</i>  |
| E-rater | - | <i>Electronic Essay Rater</i>  |
| ETS     | - | <i>Educational Testing Service</i>   |
| IR      | - | Perolehan Maklumat   |
| US      | - | <i>United States</i>   |
| MM      | - | Model <i>Multinomial</i>   |
| WWB     | - | <i>Writer's Workbench</i>  |
| AWA     | - | <i>Analytical Writing Assessment</i>   |
| VSM     | - | Model Ruang Vektor   |
| C-rater | - | <i>Criterion-rater</i>   |
| BETSY   | - | <i>Bayesian Essay Test Scoring System</i>  |
| IEMS    | - | <i>Intelligent Essay Marking System</i>  |
| SEAR    | - | <i>Schema Extract Analyse and Report</i>   |
| PS-ME   | - | <i>Paperless School free text Marking Engine</i>   |
| SVD     | - | Penguraian Nilai Tunggal   |
| LSA     | - | Analisa Semantik <i>Latent</i>   |

|           |   |  |
|-----------|---|--|
| MsNLP     | - | <i>Microsoft Natural Language Processing</i>     |
| XP        | - | <i>X Phrase</i>                                  |
| CSR       | - | Perwakilan Struktur-Konsep                       |
| Indextron | - | Rangkaian Neural Pengindeksan Corak              |
| NCAS      | - | <i>National Curriculum Assessment of Science</i> |
| ACC       | - | ketepatan keputusan                              |
| CORR      | - | kolerasi pengunduran berganda                    |
| TOEFL     | - | <i>Test of English as a Foreign Language</i>     |
| AEL       | - | purata panjang elemen hujahan                    |
| EC        | - | kandungan esei                                   |
| AC        | - | kandungan argumen                                |
| PMM       | - | Peraturan Memberi Markah                         |
| OBP       | - | Operasi Baris Permulaan                          |
| RST       | - | Teori Struktur Retorik                           |
| KB        | - | Pangkalan Pengetahuan                            |
| SD        | - | Sisihan Piawai                                   |

## SENARAI ISTILAH

|            |   |  |
|------------|---|--|
| Ambiguiti  | – | Kesamaran yang memungkinkan dua atau beberapa tafsiran.  |
| Anafora    | – | Pengulangan sesuatu kata atau frasa pada permulaan beberapa kalimat atau klausa yang berturut-turut agar memperoleh kesan tertentu.  |
| Dikotomi   | – | Pembahagian (pemisahan) antara dua kumpulan (kelompok) dalam sesuatu hal yang saling bertentangan.   |
| Diksi      | – | Pemilihan kata/gaya sebutan.   |
| Fonetik    | – | Ilmu bahasa (linguistik) yang berkaitan dengan penyebutan kata dan lambang yang menunjukkan sebutannya.  |
| Gramatis   | – | Berasaskan atau mengikut prinsip-prinsip nahu atau tatabahasa.   |
| Infleksi   | – | Penambahan imbuhan pada kata akar atau dasar.  |
| Intrinsik  | – | Sebagai sebahagian daripada sifat atau ciri seseorang atau sesuatu.  |
| Kognitif   | – | Segala perkara yang berkaitan dengan kognisi seperti proses pembelajaran, pemahaman dan pemerolehan pengetahuan.   |
| Koordinasi | – | Saling hubungan (yang dapat melicinkan perjalanan sesuatu), jalinan pertalian (tindakan, gerakan) antara bahagian-bahagian dan lain-lain yang terlibat (dalam kegiatan dan lain-lain), penyelarasan. |
| Kopula     | – | Perkataan yang menghubungkan perkara atau subjek dalam ayat dengan predikatnya.  |

|           |   |  |
|-----------|---|--|
| Leksikal  | – | Berkenaan dengan perkataan atau perbendaharaan kata sesuatu bahasa, makna sesuatu perkataan itu sendiri tanpa melihat penggunaannya dalam ayat.  |
| Leksikon  | – | Kamus, perbendaharaan kata sesuatu bahasa atau sesuatu bidang.   |
| Modus     | – | Cara, gaya, prosidur untuk melaksanakan sesuatu.   |
| Monotoni  | – | Tidak adanya keseragaman (pada nada, kerja, dan sebagainya.)   |
| Morfem    | – | Gabungan terkecil dalam bahasa yang mempunyai erti atau fungsian tertentu, contohnya perkataan ‘perumahan’ terdiri daripada satu morfem bebas iaitu perkataan ‘rumah’ dan dua morfem terikat iaitu ‘per-’ dan ‘-an’. |
| Morfologi | – | Kajian tentang pembentukan kata dalam sesuatu bahasa, termasuk infleksi, terbitan dan pemajmukan.  |
| Nahu      | – | Cabang ilmu linguistik yang berkaitan dengan sintaksis, morfologi, tatabahasa dan sebagainya.  |
| Sintaksis | – | Pengetahuan (cabang ilmu linguistik, peraturan dan sebagainya) tentang susunan kata dalam ayat.  |
| Taksonomi | – | Kajian tentang prinsip, peraturan, dan amalan dalam pengelasan organisma hidup berdasarkan persamaan dan perbezaan sifat organisma itu.  |
| Transitif | – | Kata kerja yang mempunyai penyambut (objek).   |

**SENARAI LAMPIRAN**

| <b>LAMPIRAN</b> | <b>PERKARA</b>                                 | <b>MUKA SURAT</b> |
|-----------------|--|-------------------|
| A               | Contoh Data Pengujian.                         | 158               |
| B               | Contoh Data Latihan.                           | 171               |
| C               | Skema Penilaian Esei UPSR.                     | 184               |
| D               | Skala Panduan Markah Gaya Olahan.              | 186               |
| E               | Senarai Keseluruhan Ciri Untuk Penilaian Esei. | 188               |

# **BAB 1**

## **PENGENALAN**

### **1.1 Pendahuluan**

Kebolehan komputer untuk berkomunikasi dalam bahasa tabii telah lama diperakui dalam bidang Kepintaran Buatan (AI). Kemampuannya untuk menilai sesuatu yang bersifat subjektif seperti esei untuk *Graduate Management Admissions Test* (GMAT)(Mart, 2000), telah membuka lembaran baru dalam era dunia sains dan praktikal AI. Secara tidak langsung, ianya dipercayai akan memberi impak yang besar kepada penggunaan komputer dalam lapangan pendidikan pada masa hadapan.

Sistem penilaian esei berbantuan komputer (CbAS) ini telah mula diperkenalkan kira-kira empat dekad yang lalu, iaitu pada tahun 1966 oleh Ellis Page (Page, 1994). Diinspirasikan daripada Pemprosesan Bahasa Tabii (NLP) pada masa tersebut, beberapa pengkaji di Connecticut berpendapat bahawa komputer mampu untuk memainkan peranan yang besar dalam membuat penilaian ke atas penulisan esei pelajar. Malah, sebahagian penyelidik dalam bidang NLP bersependapat bahawa CbAS dan NLP telah dirintis pada sekitar tahun 1960-an (Burstein *et al.*, 2003). Hari demi hari, para pengkaji meneruskan kajian dalam bidang ini sehingga mereka telah menempa kejayaan yang besar pada hari ini. Buktinya, beberapa buah sistem prototaip kini telah dipertingkatkan hingga ke tahap sistem pengoperasian sepenuhnya (Hearst, 2000). Bagaimanapun, ianya masih belum lagi dianggap sebagai CbAS yang cukup sempurna (Valenti *et al.*, 2003). Ini kerana, beberapa proses dalam peramalan penilaian CbAS tersebut masih boleh dipertingkatkan dan dioptimumkan pencapaiannya.



Berdasarkan kajian yang telah dilakukan oleh Rudner dan Gange (2001), terdapat tiga buah CbAS yang telah berjaya dan seringkali digunakan dalam pemarkahan esei berkomputer buat masa ini: (1) *Project Essay Grader* (PEG), diperkenalkan oleh Ellis Page pada tahun 1966; (2) *Intelligent Essay Assessor* (IEA), diperkenalkan untuk pemarkahan esei pada tahun 1997 oleh Laundauer dan Foltz (2000); dan (3) *Electronic Essay Rater* (E-rater), digunakan oleh *Educational Testing Service* (ETS) dan dibangunkan oleh Jill Burstein (Rudner dan Gange, 2001). E-rater adalah 'Teknologi Gabungan Ciri-ciri' yang mengaplikasikan kepelbagaian sintaksis, struktur hujahan (seperti PEG) dan analisa kandungan (seperti IEA). Tetapi, kajian terbaru oleh Valenti *et al.* (2003), menyatakan bahawa kini telah terdapat kira-kira 40 buah CbAS yang telah berjaya dibangunkan sebagai sistem pengoperasian sepenuhnya. Perincian tentang fakta ini akan dinyatakan dalam Bab 2: Kajian Literatur.

Setiap ujian esei dinilai berpandukan kepada skema penilaian esei yang khusus (Burstein dan Marcu, 2000). Dengan itu, pembangunan sesebuah sistem penilaian mestilah selaras dengan skema penilaian tersebut dari segi kriteria penilaian dan peratusan untuk setiap kriteria tersebut (contohnya bahagian bahasa, isi kandungan dan gaya olahan). Secara amnya, skema bagi ujian yang berlainan adalah berbeza dari segi ciri-ciri dan pembahagian markah bagi ciri tersebut. Namun begitu, masih terdapat beberapa persamaan yang wujud khususnya dari segi ciri-ciri penting yang diambil kira dalam skema penilaian tersebut. Antaranya, kesalahan bahasa, pengenalpastian elemen-elemen hujahan (pendahuluan, isi-isi penting dan penutup) dan gaya olahan (Mohd Isa, 2004). Di sini, faktor kemanusiaan memainkan peranan penting kerana ketiga-tiga ciri tersebut boleh dianggap mudah bagi manusia untuk menilainya, namun agak sukar dan menjadi masalah bagi komputer untuk memprosesnya. Namun, dengan bantuan pelbagai kemudahan seperti adanya teknik-teknik AI dan NLP serta peralatan pemrosesan yang berkemampuan tinggi, memungkinkan matlamat ini mencapai kejayaan (rujuk Jadual 2.2).

## 1.2 Latar Belakang Masalah

Esei dianggap oleh para penyelidik sebagai kayu pengukur yang sangat kondusif untuk mengukur kebolehan seseorang untuk mengingat kembali, menyusun, menjana idea dan menyatakan pendapat sendiri berbanding dengan soalan berbentuk objektif yang hanya menguji kemampuan seseorang untuk menterjemah dan mengaplikasi data semata-mata serta kurang memberikan maklumbalas yang berkesan (Gronlund, 1985). Malah, berdasarkan aras penilaian yang telah ditetapkan oleh taksonomi Bloom, penilaian esei berada di dua peringkat tertinggi dalam aras penilaian tersebut, iaitu aras penilaian dan sintesis (Bloom, 1956).

Berbanding soalan berbentuk objektif yang lebih banyak menguji kefahaman dan ingatan seseorang, soalan esei didapati lebih sukar untuk dinilai kerana ianya lebih bersifat subjektif. Jika diperhalusi, kesubjektifan ini mengundang kebarangkalian berlakunya perbezaan gred yang diberi oleh penilai yang berbeza. Malahan, penilaian yang dibuat oleh penilai yang sama ke atas esei yang sama sekalipun, tetapi pada waktu yang berbeza, berkemungkinan untuk menghasilkan gred yang berlainan. Ini berlaku disebabkan faktor emosi yang mempengaruhi tahap penilaian seseorang penilai, samada lebih ketat (*strict*), sederhana atau lebih longgar (*lenient*).

CbAS yang berkemampuan untuk menilai penulisan pelajar secara automasi, dilihat memberi impak positif berganda kepada tenaga pengajar dan mereka yang terlibat secara langsung atau tidak dalam bidang pendidikan (Palmer *et al.*, 2002). Apabila terlalu banyak esei yang perlu dinilai dalam satu-satu masa, para guru ini akan menjadi buntu dan menghadapi masalah untuk menilai secara konsisten dan memberi komen yang baik, lebih-lebih lagi jika masa yang ada terlalu singkat. Pihak pentadbir pendidikan pula menumpukan kepada komen yang berkualiti dengan masa yang minimum, tetapi dalam masa yang sama perlu mengambilkira kos yang perlu dibayar. Justeru itu, sistem penilaian ini mendapat permintaan yang tinggi sebagai alatan pendidikan, di samping dapat mengurangkan kos dan meminimakan masa penilaian (Hedberg, 1999).

Penilaian esei berautomasi ini sebenarnya telah diperolopori lebih dari 30 tahun (Williams, 2001). Namun begitu, hanya sejak akhir-akhir ini, perlaksanaannya telah dibangunkan dan diuji secara praktikal. Ini kerana, pada awal kewujudannya, sistem penilaian yang dibangunkan pada ketika itu hanya mampu mengukur kualiti esei berdasarkan ciri-ciri luaran seperti purata panjang perkataan, panjang esei, bilangan koma, bilangan kata depan dan bilangan kata retorik (Hearst, 2000). Ini secara tidak langsung menyebabkan timbulnya kritikan daripada barisan penilai dalam bidang pendidikan. Sebagai contohnya, esei yang ditulis dengan lebih panjang, akan mendapat gred yang lebih baik, tanpa mengambilkira kualiti penulisan tersebut.

Namun begitu, dengan berkembangnya penyelidikan dalam domain NLP dan Perolehan Maklumat (IR) telah memungkinan penghasilan sistem penilaian yang berkemampuan untuk mengukur bukan sahaja ciri-ciri luaran, malah yang lebih penting turut mengambilkira ciri-ciri dalaman. Page dan rakan-rakan sekolejnya, telah mendahului senarai pembangunan CbAS dengan membangunkan PEG pada tahun 1966 (Rudner dan Gange, 2001). PEG pada awal pembangunannya turut mengalami kekangan yang sama iaitu hanya mengambilkira ciri-ciri luaran, namun kini telah berupaya menilai ciri-ciri struktur (isi kandungan dan gaya olahan) untuk mengukur kualiti esei. Ianya juga merupakan implementasi penilaian esei berautomasi terawal dan paling lama bertahan.

IEA pula mengaplikasikan pendekatan "*bag of words*" dan telah dibangunkan dan dinilai oleh Landauer di University Colorado, Boulder (Tony, 2004). Ianya telah mengambilkira ciri-ciri linguistik (kesalahan bahasa) tetapi memfokuskan kepada ciri-ciri struktur untuk mengukur kualiti sesebuah penulisan.

Setelah itu, Jill Burstein telah mengorak langkah dengan membangunkan E-rater di ETS, United States (US), yang mana telah digunakan untuk menilai esei-esei GMAT pada masa kini. Model ini meningkatkan kemampuan PEG dan IEA dengan menggunakan pendekatan *hybrid* yang menggabungkan ciri-ciri linguistik (dengan mengaplikasikan teknik-teknik NLP) dan ciri-ciri struktur esei yang lain. Dengan kata lain, E-rater menggabungkan kelebihan kedua-dua model sebelum ini dengan

mengambil kira kedua-dua elemen penting iaitu ciri-ciri linguistik (bahasa) dan ciri-ciri struktur (peramalan).

Bertitik tolak dari situ, para penyelidik berlumba-lumba untuk membangunkan sebuah sistem penilaian esei berautomasi yang lebih sempurna. Kini, terdapat lebih daripada 40 buah CbAS komersial di pasaran (Valenti *et al.*, 2003). Namun begitu, ianya masih berkisar dengan isu asas sesebuah sistem penilaian yang lain iaitu sejauh mana kemampuan sistem tersebut membuat peramalan sebaik manusia. Di sini, kajian demi kajian dilakukan untuk meningkatkan tahap kesetaraan penilaian manusia-sistem agar boleh mencapai objektif utama iaitu menggantikan penilai kedua (manusia) tetapi dalam masa yang sama boleh bertindak sebaik manusia (rujuk Rajah 2.6).

Untuk tujuan itu, para penyelidik telah mengadaptasikan pelbagai teknik dan algoritma yang terbaik untuk membuat peramalan yang paling optimum ke atas kedua-dua elemen penilaian iaitu ciri-ciri linguistik dan ciri-ciri struktur. Ini kerana, kebanyakan skema penilaian esei sekarang ini mengambil kira kedua-dua elemen tersebut.

Namun begitu, sebelum sebarang teknik dapat dilaksanakan samada ke atas ciri linguistik mahupun ciri struktur, masalah pertama yang wujud ialah kegagalan sistem untuk mengenalpasti dan memperbetulkan ralat ejaan sebaik manusia (Leacock, 2004). Lebih teruk lagi, ianya turut mempengaruhi ketepatan dan peratus kesetaraan penilaian manusia-sistem di akhir pembangunan sesebuah CbAS.

Bagi kebanyakan CbAS, ianya menekankan elemen tatabahasa, perbendaharaan kata, mekanik (ejaan, imbuhan dan tanda baca), gaya penulisan (olahan) dan struktur hujahan (elemen hujahan) (Burstein dan Wolska, 2003; Leacock, 2004; Yigal, 2004; Semire Dikli, 2006; Yigal dan Burstein, 2006). Manakala sebahagiannya hanya menekankan komponen struktur hujahan, struktur sintaksis (bahasa) dan penggunaan perbendaharaan kata (Burstein *et al.*, 2001).

Burstein dan Wolska (2003) menyatakan bahawa protokol penandaan dan teknik yang berkesan diperlukan dalam menentukan dan mengelaskan elemen-elemen hujahan (penyataan tesis, isi-isi penting dan pernyataan penutup) yang wujud dalam sesebuah esei seperti pernyataan tesis dan pernyataan penutup. Mereka menerangkan bahawa walaupun penanda dilatih untuk membuat penentuan elemen hujahan, tetapi masih wujud kelemahan dari segi ketepatan pengelasan elemen hujahan tersebut. Pengelasan ini penting kerana ianya boleh digunakan oleh pelajar untuk membuat rujukan dan mengukur saling perkaitannya dengan kualiti esei yang ditulis (Burstein *et al.*, 2001). Antara teknik yang digunakan dalam menentukan dan mengelaskan argumen (ayat) dalam esei kepada sesebuah elemen hujahan yang khusus ialah Model Multinomial, *k-nearest neighbor* (pendekatan kebarangkalian), Logik Fuzzy (pendekatan berbantuan-keputusan) (Leacock, 2004). Menurut Leacock (2004) lagi, Alatan Penganalisa Penulisan sedia ada hanya menentukan elemen hujahan tetapi tidak menilai kualiti elemen tersebut. Berdasarkan kenyataan Burstein dan Marcu (2003) pula, Fuzzy dalam sesetengah keadaan menggambarkan kualiti penulisan secara umum. Namun ianya tidak menyatakan dengan terperinci ciri-ciri linguistik yang mempengaruhi kualiti tersebut.

Selain itu, masalah yang paling relevan di dalam bidang pemarkahan esei berautomasi juga ialah kesukaran untuk mendapatkan korpus esei (data) yang banyak (Christie, 2003; Larkey, 2003), setiap satunya dengan markah yang telah dinilai oleh manusia untuk tujuan latihan dan pengujian.

Dalam penyelidikan ini, untuk mengenalpasti dan mengelaskan setiap elemen hujahan dalam esei yang terdiri daripada pernyataan pendahuluan, isi-isi penting dan pernyataan penutup, kita akan mengaplikasikan Teknik Model *Multivariate Bernoulli* (MMB) daripada pendekatan Bayesian. Antara lain, penggunaan MMB ini juga adalah disebabkan ianya melibatkan saiz korpus (data) yang kecil dan mengambilkira ciri yang wujud dan tidak wujud dalam membuat penentuan dan pengelasan elemen hujahan sekaligus mengukur ciri linguistik yang mempengaruhi kualiti penulisan tersebut berbanding Logik Fuzzy dan Model *Multinomial* (MM) (Little, 2001; Burstein *et al.*, 2001; Sullivan dan Daghestani, 1997).

Bagi membuat peramalan gaya olahan, beberapa ciri telah digunakan oleh para penyelidik terdahulu. Antaranya purata panjang perkataan, tahap kegramatisan ayat, peratusan kata kerja pasif, peratusan kata nama dan penggunaan perkataan yang sama berulang kali (Burstein dan Wolska, 2003). Yi-fang Brook Wu dan Xin Chen (2005) menyatakan bahawa gaya penulisan adalah penting untuk membentuk esei berkualiti tinggi, kerana kebiasaannya, penulis akan mengelak dari menggunakan perkataan yang sama untuk menjelaskan konsep yang sama. Masalah wujud di sini ialah untuk membuat peramalan tahap gaya olahan yang lebih optimum yang menghampiri penilaian manusia menggunakan corak hubungan koefisien-pemberat dalam persamaan pengunduran yang sepadan (Valenti *et al.* 2003; Attali dan Burstein, 2006). Malah Attali dan Burstein (2006) turut menyatakan masalah penggunaan Algoritma Pengunduran Linear *Stepwise* (SLR) dalam menentukan pekali pemberat ciri yang optimum adalah kurang efisien kerana set ciri yang digunakan adalah tidak piawai di mana bilangan dan jenis ciri yang dipilih adalah berbeza untuk penilaian esei yang berlainan.

Daripada kajian yang dilakukan (Burstein *et al.*, 1998; Attali dan Burstein, 2006), CbAS semasa memerlukan pengumpulan banyak data (set esei) yang signifikan dan telah dinilai oleh manusia. Ini akan memakan masa dan memerlukan kos yang tinggi. Model terdahulu juga bergantung kepada subset pembolehubah bagi lapan hingga 12 ciri peramalan yang dipilih menggunakan Algoritma SLR daripada set yang bersaiz lebih besar, iaitu kira-kira 57 ciri (Burstein *et al.*, 1998; Burstein dan Marcu, 2000). Sebaliknya, dengan menggunakan Algoritma Pengunduran Linear Berganda (MLR), sebanyak enam ciri yang lebih optimum dan signifikan untuk membuat peramalan penilaian telah ditetapkan berdasarkan kajian-kajian penyelidik terdahulu (Attali, 2004; Attali dan Burstein, 2004; Attali dan Burstein, 2006, Semire Dikli, 2006). Dengan ini juga, ianya menjamin peramalan gaya olahan yang lebih piawai menggunakan set ciri yang bersaiz lebih kecil (Attali, 2004).

### 1.3 Pernyataan Masalah

Sepertimana telah dinyatakan dalam Subtopik 1.2, kajian ini akan menyelesaikan masalah ralat-ejaan, penentuan elemen-elemen hujahan dan peramalan gaya olahan esei.-Seterusnya, tiga pernyataan masalah yang telah dikenalpasti ialah

- (i) Bagaimanakah kaedah dan teknik kebarangkalian mampu mengenalpasti dan memperbetulkan ralat-ejaan sesebuah perkataan dalam Fasa Pra-pemprosesan dan Pengekstrakan Data?
- (ii) Sejauh manakah kebaikan pengelasan elemen-elemen hujahan yang terdiri daripada pernyataan tesis, isi-isi penting dan pernyataan kesimpulan menggunakan Teknik Model *Multivariate Bernoulli* (MMB) daripada pendekatan Bayesian menggunakan pemberat ciri tetap dapat mengenalpasti dan menganalisa Bahagian Isi Kandungan sesebuah esei?
- (iii) Bagaimanakah pemberat ciri optimum dalam Algoritma Pengunduran Linear Berganda (MLR) menggunakan set ciri bersaiz lebih kecil dapat menghasilkan persamaan pengunduran yang paling sesuai (*best fitted*) untuk memastikan peramalan gaya olahan yang lebih tepat dan seterusnya menjamin tahap kesetaraan penilaian yang lebih tinggi antara manusia dan penggunaan teknik penilaian?

### 1.4 Matlamat

Matlamat kajian ini adalah untuk mengkaji keberkesanan Teknik MMB dalam menentukan dan mengkelaskan elemen-elemen hujahan esei dan menguji kemampuan Algoritma MLR dalam membuat peramalan markah gaya olahan esei. Hasil daripada kedua-dua teknik tersebut akan diaplikasikan untuk mendapatkan gred akhir prototaip bagi membandingkan dengan gred akhir yang dinilai oleh manusia.

## 1.5 Objektif

Objektif kajian ini ialah:

- (i) Menggunakan kaedah kebarangkalian untuk mengenalpasti dan membetulkan ralat-ejaan sesebuah perkataan.
- (ii) Menggunakan pendekatan pengkelas Bayesian yang dilatih menggunakan Rumus MMB untuk menentukan dan mengelaskan elemen-elemen hujahan yang mewakili sebahagian ciri tetap peramalan esei.
- (iii) Mengaplikasikan Algoritma MLR ke atas ciri optimum untuk mendapatkan pekali pemberat paling hampir bagi menghasilkan peramalan gaya olahan yang lebih tepat.
- (iv) Menggunakan keputusan daripada objektif (ii) dan (iii) dan digabungkan dengan lima kriteria kesalahan bahasa untuk membuat perbandingan dengan penilaian manusia.

## 1.6 Skop

- (i) Esei yang dinilai hanya dalam Bahasa Melayu.
- (ii) Maksimum panjang esei yang diambil kira tidak melebihi 120 patah perkataan.
- (iii) Topik esei telah ditentukan dengan jelas.
- (iv) Menggunakan skema pemarkahan mengikut piawaian Ujian Penilaian Sekolah Rendah (UPSR) iaitu gred A, B, C, D dan E.
- (v) Hanya membangunkan model atau prototaip sistem, bukan sistem akhir.
- (vi) Format penulisan esei tidak diambil kira.
- (vii) Pembetulan ralat-ejaan hanya dibuat pada kata dasar sesebuah perkataan, tidak melibatkan imbuhan.



- (viii) 160 sampel esei digunakan untuk latihan dan 40 sampel esei digunakan untuk ujian berdasarkan prosidur *five-fold cross validation*.
- (ix) Julat kebarangkalian bagi MBM ialah -2.0 hingga 2.0
- (x) Tidak mengambilkira nilai ralat  $e$  dalam persamaan pengunduran MLR untuk tujuan pengujian gaya olahan.
- (xi) Tidak mengambilkira faktor esei yang ditulis di luar topik.
- (xii) Hanya mengambilkira bahagian isi kandungan dan gaya olahan, tetapi tidak melibatkan kriteria kesalahan bahasa.

### **1.7 Kepentingan Penyelidikan**

Kajian ini dilakukan untuk meningkatkan teknik peramalan dalam membuat penilaian esei bagi menghasilkan keputusan yang lebih baik. Semoga dengan itu juga, penilaian yang akan dilakukan lebih tepat, konsisten dan pantas berbanding manusia.

### **1.8 Sumbangan Ilmiah**

Terdapat tiga sumbangan ilmu yang dikenalpasti dari penyelidikan yang dibuat. Kedua-dua penyelidikan ini saling berkaitan di antara satu sama lain.

- (i) Pengecaman dan pembetulan ralat-ejaan pada kata dasar.
- (ii) Penentuan dan pengkelasan elemen-elemen hujahan.
- (iii) Dapatan persamaan ramalan yang lebih optimum untuk gaya olahan.

## 1.9 Struktur Tesis

Tesis ini terdiri daripada enam bab. Bab 1 terdiri daripada pengenalan kepada penyelidikan yang dilakukan. Topik-topik yang diketengahkan meliputi pengenalan kepada penyelidikan, latar belakang masalah kajian, pernyataan masalah, matlamat, objektif dan skop penyelidikan, kepentingan kajian dan sumbangan ilmiah yang diberikan.

Bab 2 membincangkan tentang kajian literatur bagi CbAS. Ianya terdiri daripada kajian terhadap sejarah awal CbAS ini yang bermula dari kajian rintis, sistem CbAS semasa beserta pencapaiannya dan teknik-teknik yang digunakan. Selain itu, turut diperjelaskan tentang pernyataan masalah penyelidikan ini dibuat dan justifikasi pemilihan teknik dan algoritma yang digunakan. Seterusnya, Skema Pemarkahan UPSR akan dihuraikan secara terperinci sebagai panduan penilaian prototaip yang dibangunkan.

Bab 3 menerangkan tentang metodologi beserta proses-proses yang terlibat dalam keempat-empat fasa iaitu Fasa I: Pra-Pemprosesan dan Pengekstrakan Ciri, Fasa II: Latihan Peramalan Gaya Olahan, Fasa III: Latihan Pengkelasan Elemen Hujahan, dan Fasa IV: Pengujian Penilaian.

Manakala Bab 4 pula memperincikan proses-proses utama bagi teknik penilaian iaitu teknik penentuan dan pengkelasan elemen hujahan (Teknik MMB dan Logik Fuzzy) dan teknik peramalan gaya olahan (Algoritma MLR dan Algoritma SLR). Selain itu, turut dibincangkan tentang pembangunan Pangkalan Data yang terlibat dan prosidur penilaian yang merangkumi rumus-rumus yang telah dinyatakan dalam Bab 3.

Bab 5 membincangkan tentang hasil pengujian menggunakan prosidur *five-fold cross validation* beserta ulasan tentang keputusan ujian berdasarkan prosidur tersebut. Satu kesimpulan untuk merumuskan pengkajian akan diketengahkan di akhir bab. Hasil dari rumusan tersebut, masalah-masalah yang timbul akan

diperjelaskan sebabnya untuk memperbaiki serta mempertingkatkan lagi teknik yang diperolehi.

Akhir sekali, Bab 6 akan membuat kesimpulan keseluruhan bagi kajian yang dilakukan beserta dengan cadangan kajian lanjutan yang mungkin berupaya untuk menambahbaik teknik-teknik yang dibangunkan.

### **1.10 Ringkasan**

Bab ini memperjelaskan justifikasi utama penyelidikan ini dilakukan. Ianya dimulai dengan membincangkan pengenalan kepada penyelidikan yang merintis pembangunan CbAS beserta contoh-contoh sistem penilaian terawal yang berjaya dibangunkan. Turut dibincangkan tentang skema pemarkahan esei yang digunakan sebagai panduan penilaian.

Latar belakang masalah kajian pula membincangkan kajian rintis CbAS yang lebih lanjut dan dari situ, masalah-masalah yang wujud pada teknik penilaian semasa dikenalpasti dan dinyatakan. Masalah-masalah tersebut diperkukuhkan dengan kajian-kajian terbaru dan memerlukan teknik serta pendekatan yang efisien untuk mengatasi masalah-masalah tersebut. Dengan itu, teknik dan pendekatan yang dipilih telah dibincangkan beserta dengan justifikasi ringkas pemilihan teknik dan pendekatan tersebut.

Kemudiannya, masalah-masalah tersebut dirumuskan sebagai pernyataan masalah dan matlamat kajian yang bemoifkan pengoptimuman teknik penilaian turut dinyatakan. Objektif kajian ini yang bertujuan untuk mengatasi masalah yang telah dikenalpasti dinyatakan dengan jelas. Manakala kekangan atau ruang lingkup kajian dijelaskan dalam skop penyelidikan. Bab ini diakhiri dengan menyatakan kepentingan kajian yang memberi motivasi untuk penyelidikan ini dilaksanakan dan sumbangan ilmiah yang diperolehi hasil daripada penyelidikan ini.

Seterusnya, dalam Bab 2, perincian perjalanan pembangunan CbAS dari kajian rintis hingga ke kajian terkini dibincangkan. Turut diperjelaskan ialah tentang beberapa CbAS terkini berserta teknik dan pencapaiannya. Selain itu, bab ini juga menerangkan masalah-masalah yang dikenalpasti dalam Bab 1 dengan lebih lanjut dan kaedah penyelesaiannya. Kaedah penyelesaian tersebut akan disokong dengan justifikasi pemilihan masing-masing. Bab 2 diakhiri dengan penerangan yang lebih mendalam tentang Skema Pemarkahan Esei UPSR yang digunakan sebagai panduan penilaian dalam penyelidikan ini.

## RUJUKAN

Abdul Aziz Abdul Talib. (1993). *Menguji Kemahiran Bahasa: Prinsip, Teknik dan Contoh*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Abu Bakar Nordin, (1986). *Asas Penilaian Pendidikan*. Petaling Jaya: Longman.

Azman Wan Chik, (1994). *Pengujian Bahasa: Kes Bahasa Melayu*, Edisi Kedua. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Aja-Fernandez, S., Alberto-Lopez, C., and Cybenko, G. V. (2002). *A fuzzy MHT Algorithm Applied to Text-Based Information Tracking*. IEE Transaction on Fuzzy Systems. vol: 10. no: 3.

Attali, Y. and Burstein, J. (2004). *Automated Essay Scoring With E-rater® V.2.0*. Conference of the International Association for Educational Assessment (IAEA). Philadelphia, PA.

Attali, Y. and Burstein, J. (2006). *Automated Essay Scoring With e-rater® V.2*. Journal of Technology, Learning, and Assessment. 4(3).

Attali, Y. (2004). *Exploring the Feedback and Revision Features of Criterion*. Paper presented at National Council on Measurement in Education (NCME). San Diego, CA.

- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I, Cognitive domain. New York, Toronto: Longmans, Green.
- Brill, E., and Mooney, R. J. (1997). *An Overview of Empirical Natural Language Processing*. American Association for Artificial Intelligence. La Canada.
- Burstein, J, Kukich, K., Wolff, S., Chi Lu, Chodorow, M., Harder, L. B., and Harris, M. D. (1998). *Automated Scoring Using A Hybrid Feature Identification Technique*. Proc. Ann. Meeting Association of Computational Linguistics. Montreal, Canada.
- Burstein, J. and Chodorow, M. (1999). *Automated Essay Scoring for Nonnative English Speakers*. Proceedings of the ACL99 Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing. College Park, MD.
- Burstein, J., Chodorow, M., and Leacock, C. (2003). *Criterion Online Essay Evaluation : An Application for Automated Evaluation of Student Essays*. Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence. Acapulco, Mexico.
- Burstein, J., Kukich, K., Wolff, S., Chi Lu., and Chodorow, M. (1998). *Enriching Automated Essay Scoring Using Discourse Marking*. Proceedings of the Workshop on Discourse Relations and Discourse Marking, Annual Meeting of the Association of Computational Linguistics. Montreal, Canada.
- Burstein, J., Kukich, K., Wolff, S., Chi Lu, and Chodorow, M. (1998). *Computer Analysis of Essays*. Proceedings of NCME Symposium on Automated Scoring. Educational Testing Service, Princeton NJ, Hunter College, New York City.

- Burstein, J., Leacock, C., and Swartz, R. (2001). *Automated Evaluation of Essay and Short Answer*. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*. Loughborough University, Loughborough, UK.
- Burstein, J., and Marcu, D. (2000). *Toward Using Text Summarization for Essay-Based Feedback*. Conferences TALN, Lausanne.
- Burstein, J., and Marcu, D. (2000). *Benefits of Modularity in an Automated Essay Scoring System*. Educational Testing Service, Princeton NJ.
- Burstein, J., Marcu, D., Andreyev, S., and Chodorow, M. (2001). *Towards Automatic Classification of Discourse Elements in Essays*. Meeting of the Association for Computational Linguistics.
- Burstein, J., Wolff, S., and Chi Lu (1999). *Using Lexical Semantic Techniques To Classify Free-Responses*. *Kluwer Academic Press*. Dordrecht, Netherlands. vol: 10.
- Burstein, J., Wolff, S., Chi Lu and Kaplan, R. M. (1997). *An Automatic Scoring System for Advanced Placement Biology Essays*. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*. Washington, D.C. pp: 174-181.
- Burstein, J., and Wolska, M., (2003). *Toward Evaluation of Writing Wtyle: Finding Overly Repetitive Word Use in Student Essays*. *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*. Budapest, Hungary. vol: 1.
- Christie, J. R. (1999). *Automated Essay Marking for Both Style and Content*. In M. Danson (Ed.), *Proceedings of the Third Annual Computer Assisted Assessment Conference*. Loughborough University, Loughborough, UK.

- Christie, J. R. (2003). *Automated Essay Marking for Content ~ does it work?*. Proceedings of the 7th CAA Conference. Loughborough, Loughborough University.
- Chodorow, M., and Leacock, C. (2000). *An Unsupervised Method for Detecting Grammatical Errors*. Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL-2000). Morgan Kaufmann, San Francisco. pp: 140-147.
- Cooper, C.R. and Odell, L. (1978). *Research on Composing: Points of Departure*. Nat'l Council of Teachers of English. Urbana, Ill.
- Cucchiarelli, A., Faggioli, E., and Velardi, P. (2000). *Will Very Large Corpora Play for Semantic Disambiguation the Role That Massive Computing Power is Playing for Other AI-hard Problems?* 2nd. Conference on Language Resources and Evaluation (LREC). Athens, Greece.
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman R. A. (1990). *Indexing by Latent Semantic Analysis*. Journal of the American Society for Information Science. 41(6). 391-407.
- De Oliveira, P. C. F., Ahmad, K., and Gillam, L.(2002). *A financial News Summarization System Based On Lexical Cohesion*. Proceedings of the International Conference on Terminology and Knowledge Engineering. Nancy, France.
- Feldman, S. (1999). *"NLP meet the Jaberwocky"*, Online, Wilton.
- Fife, B.and Berger, C. (1996). *Computer Assisted Concept Mapping and Analysis*. Paper presented at the meeting of the The National Association for Research in Science Teaching. St. Louis.



- Freedman, S. W. (1983). *Student Characteristics and Essay Test Writing Performance*. Research in the Teaching of English. vol: 17. pp: 313–324.
- Galescu, L., and Ringer, E. K. (1999). *Augmenting Words With Linguistic Information For N-gram Language Models*. Department of Computer Science, University of Rochester and NLP Group, Microsoft Research.
- Garmon, M., and Reutter, T. (1997). *The Analysis of German Separable Prefix Verbs*. Microsoft Natural Language Processing System.
- Gerdes, K., and Kahane, S. (2001). *Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy*. Proceedings of the Conference of the Association for Computational Linguistics (ACL-2001). Toulouse, France.
- Grondlund, N. E. (1985). *Mesurement and evaluation in teaching*. New York: Macmillan.
- Hamilton, R.J., Pringle, R. D., and Grant, P. M. (1992). *Syntactic Techniques for Pattern Recognition on Sampled Data Signals*. IEEE Proceedings-E. vol: 139. no. 2.
- Hearst, M. (2000). *The Debate on Automated Essay Grading*. IEEE Intelligent Systems. 15(5). 22-37. IEEE CS Press.
- Hedberg, S. R. (1999). *Computers Scoring GMAT Essays? Impossible! Or is it?*. IEEE Intelligent Systems. vol: 14. issue: 3. pp: 5-7.
- Heidorn, G. (1999). *A Handbook of Natural Language Processing Techniques*. Intelligent Writing Assistance. In: R. Dale, H. Moisl, and H. Somers (eds.). Marcel Dekker.

- Honan, W. (1999). *High Tech Comes to the Classroom: Machines that Grade Essay*. New York Times.
- Hsien-Chin Liou. (1993). *Investigation of Using Text-Critiquing Programs in a Process-Oriented Writing Class*. CALICO Journal. vol: 10. no: 4.
- Humphrey, S. M., and Shneiderman, B. (1990). *Abstract of Interest*. ACM Press, New York, NW, USA. vol: 21. issue: 4: pp: 71-78.
- Jacobs, P. (2001). *Natural Language Processing: A Brief History for Skeptics*. Unisys World. Austin.
- Jerrams-Smith, J., Soh, V., and Callear D. (2001). *Bridging Gaps in Computerized Assessment of Texts*. Proceedings of the International Conference on Advanced Learning Technologies. 139-140. IEEE.
- Kalt, T. F., and Croft, W. B. (1996). *A New Probabilistic Model of Text Classification And Retrieval*. Technical Report IR-78. University of Massachusetts Center for Intelligent Information Retrieval.
- Kogut, D. J. (2002). *Fuzzy Set Tagging*. Institute of Computer Science, Warsaw University of Technology. CICLing 2002. LNCS 2276. pp: 260-263.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Sage Publication.
- Laham, D. and Foltz, P. W. (2000). *The Intelligent Essay Assessor*. In T. K. Landauer (Ed.), IEEE Intelligent Systems.
- Landauer, T. K., Foltz, P. W., and Laham D. (1998). *An Introduction To Latent Semantic Analysis*. Discourse Processes. 25. 259-284.

- Larkey, L. S. (1998). *Automated Essay Grading Using Text Categorized Techniques*. In Proceedings of the 21st ACM/SIGIR (SIGIR-98). 90-96. ACM.
- Larkey, L. and Croft, W. B. (1996). *Combining Classifiers in Text Categorization*. Proceedings of SIGIR. 289-298.
- Leacock, C. (2004). *Scoring Free-Responses Automatically: A Case Study of a Large-Scale Assessment*. Examens. Educational Testing Service. 1(3).
- Leacock, C. (2004). *Statistical Analysis of Text in Educational Measurement*. 7th International Conference on the Textual Dated Statistical Analysis. Belgium, pp. 35-41.
- Lewis, D. D. (1992). *An Evaluation of Phrasal and Clustered Representations on A Text Categorization Task*. In Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp 37–50.
- Liang, J., and Palmer, J. D. (1994). *A Pattern Matching and Clustering Based Approach for Supporting Requirements Transformation*. Center for Software Systems Engineering. George Mason University.
- Little, J. (2001). *Computerized evaluation of essays*. Term Paper, Expert Systems.
- Lukasiewicz, J. (1930). *Philosophical remarks on many-valued systems of propositional logic*. Reprinted in Selected Works,
- L. Borowski, ed. (1997). *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam. pp. 153-179.
- Mahir Publications, (1997). *Koleksi Kertas UPSR Bahasa Malaysia..* Shah Alam: Mahir Publications Sdn. Bhd.

- MacDonald, N. H, Frase, L. T., Gingrich, P. A., and Keenan, S. A. (1982). *The Writer's Workbench: Computer Aids for Text Analysis*. IEEE Trans. Comm. Vol. COM-30. No. 1. pp. 105-110.
- Mason, O. and Grove-Stephenson, I. (2002). *Automated free text marking with paperless school*. In M. Danson (Ed.), Proceedings of the Sixth Annual Computer Assisted Assessment Conference. Loughborough University, Loughborough, UK.
- McCallum, A., and Nigam, K. (1998). *A Comparison of Event Models for Naive Bayes Text Classification*. The AAAI-98 Workshop on "Learning for Text Categorization".
- Mikko Koivisto and Kismat Sood (2004). *Exact Bayesian Structure in Bayesian Networks*. Journal of Machine Learning Research 5. 549-573.
- Milenova, B. L., Yarmus, J. S., and Campos, M. M. (2005). *SVM in oracle database 10g: removing the barriers to widespread adoption of support vector machines*. Proceedings of the 31st international conference on Very large data bases. Trondheim, Norway. pp. 1152-1163.
- Miltasakaki, E., and Kukich, K. (2000). *Automated Evaluation of Coherence in Student Essays*. Proceedings LREC-200, Linguistic Resources in Education Conf. Athens, Greece.
- Ming, P. Y., Mikhailov, A. A., and Kuan, T. L. (2000). *Intelligent essay marking system*. In C. Cheers (Ed.), Learners Together. NgeeANN Polytechnic, Singapore.
- Mitchell, T., Russel, T., Broomhead, P., and Aldridge N. (2002). *Towards robust computerized marking of free-text responses*. In M. Danson (Ed.) Proceedings of the Sixth Annual Computer Assisted Assessment Conference. Loughborough University, Loughborough, UK.

- Mohd. Isa Ab. Razak, (1988). *Soalan Subjektif: Teknik Penggubalan dan Skema Pemarkahan*. Kertas kerja. Maktab Perguruan Raja Melewar, Seremban.
- Mohd. Isa bin Abd. Razak. (2004). *Esei Pedagogi Bahasa: Pemarkahan Karangan Umum dan Karangan UPSR*. Esei Pedagogi Bahasa.
- Mokhtar Ismail. (1995). *Penilaian di Bilik Darjah*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Negnevitsky, M. (2002). *Artificial Intelligence: A guide to Intelligent Systems*. Addison-Wesley, Pearson Education. pp: 87-126.
- Nichols, P. (2005). *Evidence for the Interpretation and Use of Scores from an Automated Essay Scorer*. PEM Research Report 05-02.
- Nowson, S., Oberlander, J., and Gill, A.J. (2005). *Weblogs, Genres and Individual Differences*. In the proceedings of the 27th Annual Conference of the Cognitive Science Society. Stresa, Italy.
- Page, E. B. (1996). *Grading Essay By Computer: Why The Controversy?*. Handout for NCME Invited Symposium.
- Page, E. B. (1994). *New Computer Grading Of Student Prose, Using Modern Concepts And Software*. Journal of Experimental Education. 62(2). 127-142.
- Page, E. B., and Petersen. N. (1995). *The Computer Moves Into Essay Grading: Updating The Ancient Test*. Phi Delta Kappan. 561-565.
- Palmer, J., Williams. R., and Dreher H. (2002). *Automated Essay Grading System Applied To A First Year University Subject-How Can We Do It Better*. Proceedings of the Informing Science and IT Education (InSITE) Conference. Cork, Ireland. 1221-1229.

- Paul, N. B., Susan, T. D., and Eric, H. (2003). *Inductive Transfer For Text Classification Using Generalized Reliability Indicators*. Proceedings of the ICML-2003 Workshop on The Continuum from Labeled to Unlabeled Data. Washington DC.
- Perez, D. (2004). *Automatic Evaluation of User' S Shorts Essays by Using Statistical and Shallow Natural Language Processing Techniques*. Advanced Studies Diploma Work. Universidad Autonoma of Madrid.
- Raminah Haji Sabran, (1991). *Penilaian dan Pengujian Bahasa Malaysia, Penerapannya pada Peringkat Sekolah Rendah*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Ratnaparkhi, A. (1996). *A Maximum Entropy Part-of-Speech Tagger*. In Proceedings of the Empirical Methods in Natural Language Processing Conference. University of Pennsylvania.
- Reid, D. B. (1979). *An Algorithm for Tracking multiple Targets*. IEEE Trans. Automat. Contr. vol: AC-24. pp: 843-854.
- Robert, M. L. (2000). *Learning Syntactic Rules and Tags with Genetic Algorithms for Information Retrieval and Filtering: An Empirical Basis for Grammatical Rules*. Information Processing and Management. 32(2). pp: 185-197.
- Rudner, L. M., and Gange P. (2001). *An overview of three Approaches to Scoring Written Essays by Computer*. Assessment, Research and Evaluation, University of Maryland, College Park. vol: 7(26).
- Rudner, L. M., and Liang, T. (2002). *Automated Essay Scoring Using Bayes' Theorem*. The Journal of Technology, Learning and Assessment, Technology and Assessment Study Collaborative, Lynch School of Education. Boston College. 1(2). 3-21.

- Semire Dikli (2006). *Automated Essay Scoring*. Turkish Online Journal of Distance Education-TOJDE. vol: 7. num: 1. art: 5.
- Siegel, S., and Castellan, N. J. (1998). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.
- Siti Zanariah Satari (2003). *Multiple Linear Regression*. Universiti Teknologi Malaysia: Tesis Sarjana Muda.
- Sullivan, W. G., and Daghestani, S F. (1997). *Multivariate Analysis of Student Performance in Large Engineering Economy Classes*. Proceedings ASEE Annual Conference and Exposition. Milwaukee, WI.
- Szancewic, S., Zheng, L., Nystrom, N., and Myers, A. C. (2001). *Untrusted Hosts and Confidentiality: Secure Program Partitioning*. Proceedings of the 18th ACM Symposium on Operating System Principles (SOSP). Computer Science Department. Cornell University.
- Thompson, C. (2001). *Can Computers Understand The Meaning Of Words? Maybe, In The New On Latent Semantic Analysis*. ROB Magazine.
- Tony, T. (2004). *A Short-Answer Evaluation System*. School of Computer Science and Software Engineering. Monash University.
- Valenti, S., Cucchiarelli, A., and Panti M. (2000). *Web Based Assessment Of Student Learning*. In A. Aggarwal (Ed.). *Web-based Learning and Teaching Technologies, Opportunities and Challenges*. 175-197. Idea Group Publishing.
- Valenti, S., Cucchiarelli, A., and Panti M. (2002). *Computer Based Assessment Systems Evaluation Via The ISO9126 Quality Model*. Journal of Information Technology Education. 1(3). 157-175.

- Valenti, S., Neri, F., and Cucchiarelli, A. (2003). *An Overview of Current Research on Automated Essay Grading*. Journal of Information Technology Education. DIIGA – Università Politecnica delle Marche. Ancona, Italy. vol: 2.
- Wei Fan. (2005). *Research Track Papers: Systematic Data Selection To Mine Concept-Drifting Data Streams*. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04. Seattle, WA, USA.
- Whittington, D. and Hunt, H. (1999). *Approaches To The Computerized Assessment Of Free Text Responses*. In M. Danson (Ed.). Proceedings of the Sixth International Computer Assisted Assessment Conference. Loughborough University, UK. pp: 207-219.
- Williams, R. (2001). *Automated Essay Grading: An Evaluation Of Four Conceptual Models*. In A. Hermann and M. M. Kulski (eds). Expanding Horizons in Teaching and Learning. Proceedings of the 10th Annual Teaching and Learning Forum, Perth: Curtin University of Technology.
- Yang, Y. (1997). *An Evaluation Of Statistical Approaches To Text Categorization*. Technical Report CMU-CS-97-127. School of Computer Science. Carnegie Mellon University.
- Yi-fang Brook Wu and Xin Chen (2005). *Assessing Student Learning With Automated Text Processing Techniques*, Journal of Asynchronous Learning Network. vol: 9, issue: 3.
- Zadeh, L. A. (1965). *Fuzzy Sets*. Information and Control. 8. 338-353.



Zhou, N. (2003). *A Study on Automatic Ontology Mapping of Categorical Information*.  
Department of Geography, Land Information and Computer Graphic Facility,  
University of Wisconsin – Madison.