


**PENDEKATAN PEMBANGUNAN MODEL
REGRESI LOGISTIK MULTINOMIAL**

DIANA BINTI HASSAN



**TESIS INI DISERAHKAN UNTUK MEMENUHI
KEPERLUAN PENGIJAZAHAN IJAZAH
SARJANA SAINS**

**FAKULTI SAINS DAN SUMBER ALAM
UNIVERSITI MALAYSIA SABAH**

2017

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

**JUDUL: PENDEKATAN PEMBANGUNAN MODEL REGRESI LOGISTIK
MULTINOMIAL**

IJAZAH: SARJANA SAINS (MATEMATIK DENGAN EKONOMI)

Saya **DIANA BINTI HASSAN**, Sesi **2014-2015**, mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:

1. Tesis ini adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

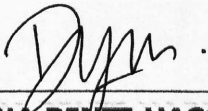
SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

TIDAK
TERHAD



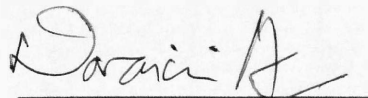
DIANA BINTI HASSAN
MS1421089T

Tarikh: 05 September 2017

Disahkan Oleh
NURULAIN BINTI ISMAIL
LIBRARIAN
UNIVERSITI MALAYSIA SABAH



(Tandatangan Pustakawan)

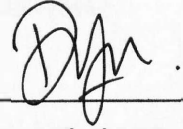


(Dr. Noraini Binti Abdullah)
Penyelia

PENGAKUAN

Saya akui karya ini adalah kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

24 Ogos 2017



Diana Binti Hassan

MS1421089T



UMS
UNIVERSITI MALAYSIA SABAH

PENGESAHAN

NAMA : **DIANA BINTI HASSAN**
NO. MATRIK : **MS1421089T**
TAJUK : **PENDEKATAN PEMBANGUNAN MODEL REGRESI LOGISTIK
MULTINOMIAL**
IJAZAH : **SARJANA SAINS (MATEMATIK DENGAN EKONOMI)**
TARIKH VIVA : **14 OGOS 2017**

DISAHKAN OLEH;

1. PENYELIA UTAMA

Dr. Noraini Binti Abdullah



Tandatangan

2. PENYELIA BERSAMA

Prof Dr. Zainodin Bin Hj. Jubok

UNIVERSITI MALAYSIA SABAH

Tandatangan

PENGHARGAAN

Segala puji ke atas Allah S.W.T, pemilik segala ilmu pengetahuan dan rahsia segenap pelusuk alam. Selawat dan salam ke atas junjungan Nabi Muhammad S.A.W yang telah diutuskan sebagaimana rahmat untuk sekalian alam, juga keluarganya, para sahabat dan sekalian yang berjuang di atas nama Ilahi.

Saya ingin merakamkan setinggi – tinggi penghargaan kepada penyelia saya iaitu Dr. Hajah Noraini Abdullah dan penyelia bersama saya iaitu Prof Dr. Zainodin Bin Haji Jubok kerana telah banyak mencurahkan ilmu yang sangat berguna sepanjang saya menyiapkan tesis ini. Kedua – dua penyelia saya ini banyak member nasihat, tunjuk ajar, doa, galakan, kesabaran dan sokongan yang padu sepanjang saya berjuang menyiapkan tesis ini. Pandangan dan pendapat yang telah dicurahkan memberikan kesan yang positif kepada keberhasilan keseluruhan kajian ini.

Setinggi-tinggi jutaan terima kasih juga kepada ahli keluarga saya yang sentiasa memberikan semangat dan dorongan dalam menyiapkan tesis ini, rakan – rakan seperjuangan yang banyak memberikan tunjuk ajar dan memperbetulkan kesilapan yang saya terlepas pandang dan semua orang yang berada disekeliling saya sepanjang perjuangan saya menyiapkan tesis ini.

Akhir sekali, ucapan terima kasih kepada para pensyarah Matematik Dengan Ekonomi, pensyarah Fakulti Sains dan Sumber Alam dan kakitangan Pusat Pengajian Pascasiswazah yang telah memberikan sokongan secara langsung mahupun tidak dan semua pihak yang turut terlibat sehingga jaya berjaya menyiapkan penulisan tesis ini.

Diana Binti Hassan

24 Ogos 2017

ABSTRAK

Analisis data yang mempunyai pembolehubah bersandar berbentuk kualitatif melibatkan lebih daripada dua kategori dikenali sebagai Regresi Logistik Multinomial. Kajian ini bertujuan untuk membina pendekatan pembangunan model Regresi Logistik Multinomial. Model ini merupakan kesinambungan daripada model Regresi Berganda dan Binari Logit Berganda. Selain memperkenalkan pembolehubah tak bersandar berbentuk patung (*dummy*), kombinasi pembolehubah patung interaksi tahap pertama turut dibuktikan dalam pendekatan ini. Prosedur ini menggunakan penyingkiran pembolehubah tidak signifikan secara progresif dalam fasa kedua yang melibatkan tiga jenis ujian iaitu ujian NPC/NPM, ujian Multikolinearan dan ujian Pekali. Penggunaan *Pseudo R-Square* yang terdiri daripada Cox & Snell, Nagelkerke dan McFadden dicadangkan sebagai penentu dalam membuat pemilihan model terbaik. Prosedur yang terpilih telah dikupas, ditunjukkan dan diterangkan secara terperinci dalam tesis ini. Penemuan yang ditonjolkan dalam kajian ini adalah pengesahan model terbaik yang dilakukan dengan menggunakan Peratusan Ralat Min Mutlak (MAPE) dan kaedah penganggaran data yang hilang. Kaedah penganggaran data hilang telah dikaji untuk memastikan model terbaik yang diperolehi memberi penganggaran nilai; pembolehubah bersandar dan ramalan yang lebih baik. Menerusi model terbaik daripada setiap kategori didapati bahawa pendekatan pembangunan model Regresi Logistik Multinomial telah diperkukuhkan. Selain itu, pembolehubah interaksi patung turut memberikan sumbangan dalam model terbaik bagi semua kategori pembolehubah bersandar.

Kata Kunci: Regresi Logistik Multinomial, Kategori Pembolehubah Bersandar, Pendekatan Pembangunan Model, *Pseudo R-Square*, MAPE, Penganggaran Data Hilang

ABSTRACT

THE MODEL-BUILDING APPROACH IN MULTINOMIAL LOGISTIC REGRESSION

Data analysis with qualitative dependent variables involves more than two categories which is known as Multinomial Logistic Regression. This research is aimed to develop a procedure for Multinomial Logistic Regression analysis. This model is a continuation of a Multiples regression model and Binary Logistic Regression. Besides introducing the independent variables in the form of dummies, a combination of first order interaction dummy variables are also included. A progressive elimination (one by one, least significant first) of the insignificant variables is employed at phase 2 involving three types of tests namely NPC / NPM, multicollinearity and coefficient test. It was also proposed to use pseudo R-square (Cox & Snell, Nagelkerke and McFadden) to finally single out the best model. Detailed procedure was explored, illustrated and explained in the thesis. The important findings highlighted in this study are the best model validation which was performed using Mean Absolute Percentage Error (MAPE) and the method of how to estimate the missing values. An extended work involving estimation of missing values are included in this work to ensure that he best model obtained would give a better estimate and forecast. Through the best models from each category, the model-building approach in Multinomial Logistic Regression analysis is established. In addition, it is found that an interaction dummy variable also contributes in the best model for all categories of the dependent variable.

Keyword: Multinomial Logistic Regression, Categorical Dependent Variable, Model Building Procedure, Pseudo R-Square, MAPE, Estimate Missing Value

SENARAI KANDUNGAN

Halaman

TAJUK	i
PENGAKUAN	ii
PENGESAHAN	iii
PENGHARGAAN	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
SENARAI KANDUNGAN	vii
SENARAI JADUAL	x
SENARAI RAJAH	xii
SENARAI SIMBOL	xiii
SENARAI SINGKATAN	xiv
SENARAI LAMPIRAN	xv
BAB 1: PENDAHULUAN	1
1.1 Pengenalan	1
1.2 Penyataan Masalah	2
1.3 Objektif Kajian	3
1.4 Skop Kajian	3
1.5 Kepentingan Kajian	4
BAB 2: ULASAN LITERATUR	6
2.1 Regresi Logistik	6
2.2 Regresi Logistik Multinomial	8
2.3 Kajian Lepas Pendekatan Pembangunan Model	11
2.4 Kajian Lepas Ujian Statistik : Ujian Multikolinearan	12
2.5 Kajian Kes : Industri Perfileman	15
BAB 3: METODOLOGI	17
3.1 Pengenalan	17
3.2 Pengumpulan Data	18
3.3 Pembahagian Data	18

3.4	Analisis Faktor	18
3.5	Penjelmaan Pembolehubah	21
3.6	Pendekatan Pembangunan Model	21
3.6.1	Fasa 1: Semua Model Berkemungkinan	22
3.6.2	Fasa 2: Model Terpilih	26
	a) Ujian Multikolinearan	26
	b) Ujian Pekali	30
3.6.3	Fasa 3: Pemilihan Model Terbaik	37
	a) Pseudo-R-Squared (R^2)	37
	i) Cox and Snell (R^2)	37
	ii) Nagelkerke (R^2)	38
	iii) McFadden (R^2)	38
3.6.4	Kebagusan Penyuaian Model	38
	a) Pearson Khi-Kuasa Dua dan Ujian Devians	39
3.7	Peratusan Ralat Min Mutlak (<i>Mean Absolute Percentage Error</i>)	40
BAB 4: ANALISIS DAN KEPUTUSAN		42
4.1	Analisis Regresi Logistik Multinomial	42
4.2	Bahagian 1: Pendekatan Pembangunan Model Tanpa Interaksi	42
4.2.1	Fasa 1: Semua Model Berkemungkinan untuk Setiap Kategori	42
4.2.2	Fasa 2: Model Terpilih	43
	a) Fasa 2.0: Ujian R^2	44
	b) Fasa 2.1: Ujian Multikolinearan	45
	c) Fasa 2.2: Ujian Pekali	46
4.2.3	Fasa 3: Model Terbaik	53
4.2.4	Fasa 4: Kebagusan Penyuaian	59
4.3	Bahagian 1: Pendekatan Pembangunan Model Dengan Interaksi	64
4.3.1	Fasa 1: Semua Model Berkemungkinan	64
4.3.2	Fasa 2: Model Terpilih	65
	a) Fasa 2.0: Ujian R^2	65
	b) Fasa 2.1: Ujian Multikolinearan	67
	c) Fasa 2.2: Ujian Pekali	67
4.3.3	Fasa 3: Model Terbaik	68

4.3.4 Fasa 4: Kebagusan Penyuaian	76
4.4 Bahagian 2: Peratusan Ralat Min Mutlak (MAPE)	78
4.5 Bahagian 3: Penganggaran Data Tak Tercerap	79
BAB 5: PERBINCANGAN DAN KESIMPULAN	80
5.1 Perbincangan	80
5.2 Pencapaian Objektif Kajian	80
5.3 Kekangan Kajian	82
5.4 Cadangan Di Masa Hadapan	83
RUJUKAN	85
LAMPIRAN	92

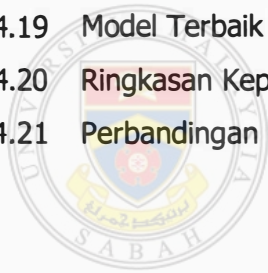


UMS
UNIVERSITI MALAYSIA SABAH

SENARAI JADUAL

	Halaman
Jadual 3.1 <i>Rotated Component Matrix</i>	20
Jadual 3.2 Ringkasan Taburan Kebarangkalian, Y_i	25
Jadual 3.2 Tafsiran umum nilai MAPE	41
Jadual 4.1 Semua Model Berkemungkinan`	43
Jadual 4.2 Jumlah Model Berkemungkinan	43
Jadual 4.3 Senarai Nilai R^2 bagi Model M13 kategori 1	45
Jadual 4.4 Pembuangan Sumber Multikolinearan daripada model M13 Kategori 1	46
Jadual 4.5(a) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.0 Kategori 1	47
Jadual 4.5(b) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.1 Kategori 1	48
Jadual 4.5(c) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.2 Kategori 1	48
Jadual 4.5(d) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.3 Kategori 1	49
Jadual 4.5(e) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.4 Kategori 1	50
Jadual 4.5(f) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.5 Kategori 1	50
Jadual 4.5(g) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.6 Kategori 1	51
Jadual 4.5(h) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.7 Kategori 1	51
Jadual 4.5(i) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.8 Kategori 1	52
Jadual 4.5(j) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.9 Kategori 1	52
Jadual 4.5(k) Penyingkiran Pembolehubah yang Tidak Signifikan bagi Model M13.0.10 Kategori 1	52
Jadual 4.5(l) Penyingkiran Pembolehubah yang Tidak Signifikan bagi	

	Model M13.0.11 Kategori 1	53
Jadual 4.6	Nilai yang Berkaitan dengan Kriteria Pemilihan Bagi Setiap Model Terpilih Dalam Kategori 1	53
Jadual 4.7	Model Terbaik Bagi Setiap Kategori	58
Jadual 4.8	Kebagusan Penyuaian Model Kategori 1	60
Jadual 4.9	Ujian Nisbah Kebarangkalian Kategori 1	61
Jadual 4.10	Kebagusan Penyuaian Model Kategori 2	61
Jadual 4.11	Ujian Nisbah Kebarangkalian Kategori 2	62
Jadual 4.12	Kebagusan Penyuaian Model Kategori 3	63
Jadual 4.13	Ujian Nisbah Kebarangkalian Kategori 3	64
Jadual 4.14	Nilai R^2 Model M338 Kategori 1	66
Jadual 4.15	Fasa 2.0 Ringkasan Model M338	66
Jadual 4.16	Fasa 2.1 Ringkasan Model M338	67
Jadual 4.17	Fasa 2.1 Ringkasan Model M338	68
Jadual 4.18	Model Terpilih Kategori 1	68
Jadual 4.19	Model Terbaik Interaksi Tahap Pertama Bagi Setiap Kategori	73
Jadual 4.20	Ringkasan Keputusan MAPE bagi Setiap kategori	79
Jadual 4.21	Perbandingan Nilai Sebenar dan Nilai Anggaran Data Tak Tercerap	79



UNIVERSITI MALAYSIA SABAH

SENARAI RAJAH

		Halaman
Rajah 3.1	Kerangka Kerja Penyelesaian Permasalahan Kajian	17
Rajah 3.2	<i>Scree Plot</i>	19
Rajah 3.3	Prosedur Pembangunan Model	21



UMS
UNIVERSITI MALAYSIA SABAH

SENARAI SIMBOL

α	Alpha (Tahap Keertian Dalam Ujian Hipotesis)
β	Pekali
Ω_0	Pemalar
Ω_j	Pekali ke-j untuk pembolehubah tidak bersandar
u	Ralat
k	Bilangan Pembolehubah Tak Bersandar
q	Bilangan Pembolehubah Tunggal Tak Bersandar
L	Fungsi Kebolehjadian
LL	Log-Kebolehjadiannya
n	Saiz Sampel
q	Bilangan Semua Model Berkemungkinan
H_0	Hipotesis Nol
H_1	Hipotesis Alternatif
$=$	Sama Dengan
\neq	Tidak Sama Dengan
$<$	Lebih Kecil Daripada
$>$	Lebih Besar Daripada
\leq	Kurang Daripada atau Sama Dengan
\geq	Lebih Kurang atau Sama Dengan
$\%$	Peratus
Σ	Sigma (Jumlah Hasil Tambah)
π	Pi (Pendaraban)
R^2	Pekali Penentuan
σ^2	Varians
$ $	Modulus
Se	Ralat Piawaian
df	Darjah Kebebasan
μ	Populasi Min
ρ	Rho (Pekali Korelasi)
ΔG^2	Ujian Statistik Nisbah Kebolehjadian

SENARAI SINGKATAN

APM	<i>All Possible Models</i>	Semua Model Berkemungkinan
MAPE	<i>Mean Absolute Percentage Error</i>	Peratusan Ralat Min Mutlak
MBL	<i>Multiple Binary Logistic</i>	Logistik Binari Berganda
MLR	<i>Multinomial Logistic Regression</i>	Regresi Logistik Multinomial
MR	<i>Multiple Regression</i>	Regresi Berganda
MS	<i>Mean Square</i>	Min Kuasa-Dua
N	<i>Number of Observations</i>	Bilangan Pemerhatian
NP	<i>Number of Parameter</i>	Bilangan Parameter
OLS	<i>Ordinary Least Square</i>	Kaedah KuasaduaTerkecil Biasa
SPSS	<i>Statistical Package for the Social Science</i>	Pakej Statistik Bagi Sains Sosial
SSE	<i>Sum of Squared Errors</i>	Jumlah Kuasa-Dua Ralat
SST	<i>Total Sum of Squares</i>	Jumlah Hasil Tambah KuasaDua
VIF	<i>Variance Inflation Factor</i>	Faktor Inflasi Varians
8SC	Eight Selection Criteria	Lapan Kriteria Pemilihan



UMS
UNIVERSITI MALAYSIA SABAH

SENARAI LAMPIRAN

Lampiran A

92

Lampiran B

105



UMS
UNIVERSITI MALAYSIA SABAH

BAB 1

PENDAHULUAN

1.1 Pengenalan

Analisis regresi merupakan kaedah statistik yang digunakan untuk mengkaji hubungan antara pembolehubah-pembolehubah. Ia telah digunakan dalam pelbagai bidang seperti sains sosial, pentadbiran, pertanian, perubatan dan pendidikan. Analisis regresi mula dibangunkan oleh Francis Galton pada akhir abad ke-19 (Kutner *et al.*, 2005). Galton telah mengkaji hubungan antara ketinggian anak dengan ibu bapa. Hasilnya, ketinggian anak adalah mengikuti ketinggian ibu bapa. Maka, Galton membangunkan penerangan matematik untuk menunjukkan kecenderungan regresi dan menjadi pelopor model regresi hari ini. Antara hubungan lain yang wujud dan dikaji ialah hubungan antara perbelanjaan dengan bilangan ahli dalam keluarga, hubungan antara kadar tindak balas kimia dengan suhu, hubungan antara hasil tanaman dengan faktor cuaca, hubungan antara pencapaian pelajar SPM dengan kelayakan masuk ke peringkat asasi dan hubungan antara kadar kesesakan jalan raya dengan bilangan kenderaan di jalan raya.

Terdapat pelbagai jenis teknik regresi yang wujud untuk membuat ramalan. Teknik-teknik ini kebanyakannya didorong oleh tiga metrik iaitu bilangan pembolehubah tak bersandar, jenis pembolehubah bersandar dan bentuk garisan regresi. Antaranya ialah regresi linear sederhana, linear berganda, logistik (binari dan multinomial) dan polinomial. Pada asasnya, model regresi adalah salah satu cara yang formal untuk menyatakan hubungan penting dalam statistik iaitu kecenderungan pembolehubah bersandar (Y) untuk mengalami perubahan kesan daripada pembolehubah tak bersandar (X) dengan cara yang sistematik; dan yang kedua hubungan statistik melalui titik yang berselerak sekitar lengkungan. Kedua-dua ciri tersebut menunjukkan terdapat taburan kebarangkalian Y bagi setiap tahap X. Tujuan analisis regresi adalah sebagai satu penerangan, penganggaran dan ramalan (Rudolf *et al.*, 2006). Data yang diambil daripada setiap pembolehubah

digunakan untuk melihat bentuk hubungan atau model yang wujud, sama ada hubungan linear atau tidak linear dan penganggaran kesemua parameter dalam model tersebut.

1.2 Penyataan Masalah

Kajian berkaitan pendekatan pembangunan model untuk mencari model terbaik khususnya bagi regresi logistik Multinomial belum ada diperkenalkan. Model terbaik boleh digunakan untuk meramal nilai pembolehubah tunggal (Zainodin dan Khuneswari, 2009). Pendekatan yang telah ada dan boleh digunakan adalah bagi regresi berganda dan regresi logistik binari berganda sahaja. Hal ini dapat diperhatikan bahawa model terbaik regresi berganda bagi penentuan harga penjualan rumah berjaya diperolehi dan boleh digunakan untuk meramal serta membuat keputusan yang logik (Zainodin dan Khuneswari, 2009). Selanjutnya, model terbaik yang diperolehi dalam analisis regresi logistik binari berganda telah memperkukuhkan pendekatan pembinaan model dimana pembolehubah interaksi peringkat-dua adalah signifikan dalam model terbaik bagi semua kajian kes (Zainodin dan Khuneswari, 2010).

Namun begitu, kajian-kajian tersebut melibatkan pembolehubah bersandar berbentuk kuantitatif bagi regresi berganda dan pembolehubah bersandar bersifat kualitatif dua kategori yang mengambil nilai 0 atau 1 bagi regresi logistik binari berganda. Walau bagaimanapun, kajian yang melibatkan pembolehubah bersandar bersifat kualitatif lebih daripada dua kategori tidak sesuai dengan pendekatan yang sedia ada untuk meneliti model terbaik. Maka pendekatan yang sesuai diperlukan bagi menyelesaikan kajian kes.

Dielman (2001) menganalisis semua pembolehubah tak bersandar yang berpotensi menyumbang kepada pembolehubah bersandar dan mencari model terbaik dimana pekali pembolehubah tak bersandar yang terlibat seharusnya signifikan. Maka, semua model yang berkemungkinan perlu diuji menggunakan beberapa ujian yang berkaitan bagi memudahkan proses pemilihan. Ini kerana, pengkaji berusaha untuk memasukkan kesemua pembolehubah tak bersandar yang berkemungkinan dalam analisis untuk mengelakkan maklumat yang penting tersingkir atau hilang.

Dalam kehidupan seharian, situasi pembolehubah bersandar dipengaruhi oleh banyak pembolehubah tak bersandar sering berlaku. Apabila berhadapan dengan situasi tersebut, masalah keraguan akan wujud sama ada hendak menggunakan kesemua pembolehubah tersebut atau buang beberapa pembolehubah tak bersandar. Jika pembolehubah tak bersandar yang berlebihan dibuang, wujud masalah lagi sama ada betul pembolehubah yang dibuang tidak akan menyumbang kepada pembolehubah bersandar.

Kajian kes iaitu pemasaran filem di Malaysia digunakan untuk mendapatkan gambaran jelas pendekatan bagi mencari model terbaik regresi logistik Multinomial. Data kajian kes memiliki bentuk pembolehubah yang bersesuaian dengan model. Tambahan, cabaran yang dihadapi oleh industri perfileman berkaitan penjualan tiket yang tidak menentu memerlukan penelitian pelan jangka masa panjang (Harian Metro, 2017). Faktor-faktor yang mempengaruhi ekosistem penonton di Malaysia perlu diambil perhatian pihak penerbit semasa memperluaskan jaringan promosi dan pemasaran baru. Maka, ramalan faktor-faktor yang signifikan mempengaruhi penonton memerlukan kajian lanjut bagi membantu pengusaha filem menghasilkan filem yang baik dan diterima penonton secara amnya di Malaysia (Harian Metro, 2017).

1.3 Objektif Kajian

Objektif kajian ini adalah bertujuan:

- i. Untuk membina pendekatan pembangunan model dalam memperoleh model terbaik regresi logistik Multinomial (MLR).
- ii. Untuk mengenalpasti pembolehubah yang signifikan pada interaksi tertinggi tahap pertama yang mempengaruhi pembolehubah bersandar.
- iii. Untuk menganggarkan nilai pembolehubah tak bersandar yang hilang menggunakan model terbaik.

1.4 Skop Kajian

Dalam kajian ini, prosedur untuk memperoleh model terbaik regresi logistik Multinomial akan dibina dengan menggunakan pembolehubah tak bersandar dan pembolehubah bersandar berbentuk kategori. Parameter akan dianggarkan menggunakan keadah penganggaran *maximum likelihood* (ML). Gambaran yang

jelas berkaitan prosedur dan analisis model regresi logistik Multinomial (MLR) akan diilustrasikan menggunakan kajian kes dalam bidang perfileman. Data kajian ini mempunyai saiz sampel sebanyak 1337 ($n=1337$) yang melibatkan beberapa ibu negeri di Malaysia iaitu Sabah, Sarawak, Johor, Selangor, Pahang dan Kedah. Data diperolehi dengan mengedarkan borang soal selidik kepada pengunjung di pawagam yang terlibat.

Semua prosedur yang terlibat seperti mengenalpasti pembolehubah, analisis faktor sehingga proses mendapat persamaan kekerapan penontonan menggunakan model regresi logistik Multinomial (MLR) akan diperjelaskan dalam kajian ini. Fokus kajian ini adalah untuk mendapatkan model terbaik logistik Multinomial yang bebas daripada masalah multikolinearan dan pembolehubah yang tidak signifikan menggunakan pendekatan pembangunan model.

1.5 Kepentingan Kajian

Pendekatan pembangunan model adalah penting untuk mendapatkan model terbaik yang akan menerangkan gambaran sebenar data set yang digunakan. Pendekatan ini boleh digunakan dalam pelbagai bidang lain seperti ekonomi, sains persekitaran, sains hayat dan perubatan. Pendekatan ini menumpukan kepada semua model yang berkemungkinan sehingga interaksi pembolehubah yang tertinggi tahap pertama dan penyingkiran pembolehubah yang mempunyai masalah multikolinearan dan tidak signifikan daripada semua model yang berkemungkinan.

Menurut Randall (2005), multikolinearan wujud dalam model ramalan apabila dua atau lebih pembolehubah dalam model adalah berkait antara satu sama lain. Ia mendorong kepada ketidakupayaan dalam menganalisis kepentingan pembolehubah input dalam model berdasarkan kepada ujian statistik yang tidak sah, dimana, pembolehubah input yang menyediakan maklumat berlebihan (bertindih) atau sama dalam menjelaskan dan meramalkan kesan yang mendasari pembolehubah bersandar dalam model. Multikolinearan berlaku apabila wujud korelasi yang sangat tinggi antara pembolehubah tak bersandar yang boleh menyebabkan nilai anggaran parameter jauh berbeza dengan nilai sebenar parameter tersebut (Warner, 2008). Justeru, masalah multikolinearan perlu diselesaikan ke atas data terlebih dahulu dan bukan dengan model ramalan. Tanpa

masalah multikolinearan atau dengan kesilapan piawai yang lebih rendah, pekali yang lain mungkin menjadi signifikan.

Jackson (2012) juga membincangkan penglibatan pembolehubah yang tidak signifikan ke dalam persamaan. Kesannya, pembolehubah tersebut cenderung untuk meningkatkan anggaran ralat piawai dan boleh menyebabkan pembolehubah yang signifikan menjadi tidak signifikan. Idea tersebut turut disokong oleh Haines dan Flori (2013), dimana penyikiran pembolehubah tak bersandar yang tidak signifikan daripada model boleh membantu mengurangkan kedimensian model dan meningkatkan tahap ketepatan anggaran ralat piawai.

Ramalan faktor-faktor signifikan bagi kajian kes amat penting bagi membantu FINAS dalam mengalakkan penerbitan filem berkualiti yang mampu menarik perhatian dan motivasi penonton ke pawagam. Ini kerana, kejayaan sesebuah filem datangnya daripada kualiti filem yang dihasilkan. Pihak penerbit perlu menerbitkan filem yang bukan sahaja diterima baik di dalam negara malah mampu menjadi komoditi eksport ke luar negara. Justeru, perhatian ke atas faktor yang mampu mendidik penonton untuk menyokong filem yang memberikan nilai tinggi sinema dapat dibuktikan menggunakan model terbaik yang diperolehi daripada regresi logistik Multinomial.

BAB 2

ULASAN LITERATUR

2.1 Regresi Logistik

Regresi logistik memiliki pembolehubah bersandar berbentuk kategori manakala pembolehubah tak bersandar berbentuk kuantitatif, kualitatif atau perduaan. Ia sesuai digunakan untuk mendapatkan model penerangan yang tepat (Zainodin dan Khuneswari (2008) dan merupakan satu analisis yang digunakan untuk menganggarkan keahlian mutlak pembolehubah bersandar tanpa mengambil kira andaian yang berkaitan dengan pembolehubah tak bersandar (Camdeviren *et al.*, 2007). Ini banyak digunakan dalam aplikasi pembelajaran mesin, sains komputer (struktur data), botani (klasifikasi) dan psikologi (teori keputusan) serta sebagai model ramalan dalam bidang perubatan. Kini, model regresi logistik kerap digunakan dalam penyelidikan perubatan bagi menentukan faktor-faktor yang berisiko serta proses mendiagnosis pesakit. Hasil kajian yang mempunyai pembolehubah bersandar berbentuk perduaan dikenali sebagai regresi logistik binari (Lawrence *et al.*, 2013).

Regresi logistik menggabungkan dua pendekatan statistik yang sangat berbeza (SPSS, 2009). Pertama ialah analisis jadual kontingensi iaitu merentasi tabulasi atau *crosstabs*, dimana kesemua pembolehubah diukur sama ada dikotomi (dua kategori), nominal (lebih daripada dua kategori dengan tidak wujud susunan antara kategori) atau ordinal (lebih daripada dua kategori dengan wujud beberapa susunan lebih besar daripada atau kurang daripada kategori) dan kebiasaannya pembolehubah mempunyai sedikit kaitan antara satu sama lain lazimnya kurang daripada 10 dan bagi kategori yang berbeza selalunya kurang daripada 20. Berkembang daripada dua hala atau jadual kontingensi bivariat (dua pembolehubah) kepada tiga hala atau trivariat (tiga pembolehubah) dengan pelbagai cara atau multivariat (pelbagai pembolehubah), jadual kontingensi membawa kepada perkembangan *log-linear* untuk berurusan dengan pelbagai

dimensi jadual kontingensi yang kompleks.

Pendekatan yang kedua ialah kuasadua terkecil biasa (*ordinary least square*, OLS) analisis regresi berganda, dimana pembolehubah biasanya diukur pada tahap nisbah atau digabungkan sebagai pembolehubah patung dimana satu set untuk beberapa pembolehubah dikotomi yang digunakan untuk mewakili pembolehubah pelbagai kategori, kebiasaannya pembolehubah nominal (SPSS, 2009). Contohnya pembolehubah etnik yang diwakili dengan mengasingkan pembolehubah-pembolehubah untuk Amerika, Africa, Hispanic atau Latin dan lain-lain. Teknik-teknik asal yang digunakan dalam analisis *log-linear* tidak menangani pengukuran ramalan pada jarak berterusan atau tahap skala nisbah dengan baik dan OLS regresi tidak mengendalikan pembolehubah bersandar dikotomi, nominal atau ordinal dengan baik. Oleh itu, model regresi logistik menunjukkan keberangkalian kehadiran atau ketidak hadiran nilai cerapan pembolehubah bersandar yang diberikan.

Penggunaan logistik ini telah memuncak sejak sedekad yang lalu bermula dari bidang kajian biomedik dan berkembang ke bidang lain seperti perniagaan dan kewangan, kriminologi, ekologi, kejuruteraan, polisi perubatan dan hidupan liar biologi (Hosmer and Lameshow, 2000). Ini kerana ketersediaan yang wujud dalam pakej perisian statistik berserta dengan prosedur. Regresi logistik telah menjadi kegemaran ramai seperti yang telah dijangka oleh Huck (2004) yang meramal regresi logistik akan mengambil alih regresi berganda dan menjadi alat regresi yang paling kerap digunakan.

Jika disoroti kajian yang terdahulu didapati bahawa Ozdemir (2011) telah menggunakan kaedah regresi logistik dalam persekitaran Sistem Maklumat Geografi (GIS) bagi menghasilkan peta untuk mencari zon yang berpotensi bagi mata air bawah tanah di Gunung Sultan tengah Turki. Menerusi dapatan kajian pengkaji, model yang diperolehi daripada regresi logistik menjadi penganggar yang baik berpotensi pada musim bunga di kawasan kajian. Tambahan, model boleh dikembangkan untuk meningkatkan kesesuaian data air bawah tanah musim bunga yang ada. Kajian ini mencadangkan kaedah ini di gunakan secara rutin dalam penerokaan air bawah tanah bagi keadaan yang menggalakkan.

Disamping itu, Young dan Lisman (2007) turut membangunkan model kuantitatif yang mengaitkan rekod kemalangan trak terbalik di Wyoming bagi mengukur kelajuan angin di stesen cuaca yang berhampiran. Model logistik binari telah dianggarkan daripada data untuk menentukan jika ada korelasi yang signifikan diantara data angin stesen cuaca dengan kemungkinan berlakunya kemalangan kenderaan terbalik. Menerusi dapatan kajian mereka, data stesen cuaca boleh digunakan sebagai peramal kemalangan kenderaan terbalik. Kajian ini juga menjadi langkah permulaan yang diperlukan untuk pembangunan peraturan operasi bagi jalan raya yang berisiko tinggi menyebabkan kemalangan trak terbalik dalam keadaan kelajuan angin yang tinggi.

Antara kajian lain yang telah menggunakan kaedah regresi logistik termasuklah ramalan kilat di Pusat Angkasa Kennedy dan Stesen Tentera Udara Cape Canaveral (Lambert, 2007; Lambert dan Wheeler 2005), bencana tanah runtuh di Kansas (Ohlmacher dan Davis, 2003; Elward *et al.*, 2002), sikap pemeliharaan alam sekitar di Nepal (Mehta dan Heinen 2001), pengesanan memfailkan penyata kewangan palsu di Greece (Spathis, 2002), penggunaan lembaga pengarah luar oleh firma-firma swasta kecil di Amerika Syarikat (Fiegner *et al.*, 2000), pengurusan perubatan sendiri dalam kalangan dewasa yang lebih tua (Madigan *et al.*, 2003), kadar kegagalan pelajar tahun pertama (Patrick, 2001), pilihan berjalan untuk pengangkutan dan rekreasi (Lee dan Moudon, 2006) dan pelanggaran kenderaan dengan hidupan liar di Pergunungan Berbatu Kanada (Gunson *et al.*, 2003). Justeru, regresi logistik merupakan teknik statistik yang boleh diaplikasi dalam pelbagai bidang.

2.2 Regresi Logistik Multinomial

Regresi logistik Multinomial (MLR) ialah kaedah berkategori dalam statistik yang menjadikan regresi logistik binari kepada pelbagai kategori masalah, contohnya lebih daripada dua hasil yang berkemungkinan. Maka, MLR ialah model yang digunakan untuk meramal keberangkalian perbezaan hasil kategori yang berkemungkinan sebagai taburan pembolehubah bersandar dengan diberikan satu set pembolehubah tak bersandar (nilai nyata, nilai binari dan nilai mutlak). MLR juga dikenali dengan pelbagai nama seperti pengelasan regresi linear, regresi