

Kajian Tingkah Laku Pengguna Berdasarkan Gambar Profil Media Sosial (*Instagram*)
(*Study of User Behavior Based on Image Profile Used in Social Media (Instagram)*)

SITI UBAIDAH RAMLAN, SUZAIMAH RAMLI*, SYAHANEIM MARZUKHI

ABSTRAK

Kandungan imej yang disiarkan pengguna ke media sosial didorong sebahagiannya oleh personaliti mereka. *Instagram* adalah tempat untuk mendedahkan dan perkongsian cara hidup dan gaya diri. Memberi pendedahan perkongsian diri terlalu banyak di *Instagram* boleh membentuk identiti yang berbeza antara di dunia nyata dan dalam talian. Penyelidikan ini dijalankan pada 100 pengguna di Universiti Pertahanan Nasional Malaysia. Objektif kajian ini adalah untuk mengkaji dengan lebih teliti mengenai kelakuan, perwatakan, tingkah laku, serta personaliti seseorang di media sosial. Pengguna media sosial mudah terpedaya dengan apa yang sering ditonjolkan di media sosial. Oleh itu kajian ini menyumbang kepada ketelitian terhadap tingkah laku personaliti berdasarkan gambar profil media sosial masing-masing. Kajian tingkah laku pengguna terhadap gambar profil media sosial menggunakan metodologi *CRIPS-DM* iaitu *Cross Industry Standard Process for Data Mining*. Perbandingan di bahagian ini bersesuaian dengan menggunakan algoritma *Random Forest* dan *Decision Tree* dalam bahagian *Classification*.

ABSTRACT

The image content that users post to social media is driven in part by their personalities. Instagram is a place to reveal and share your lifestyle and style. Giving too much self-sharing exposure on Instagram can form a different identity between the real world and online. This research was conducted on 100 Instagram users at Universiti Pertahanan Nasional Malaysia. With this research project, it can look carefully at the behavior, character, behavior, personality of a person on social media and easily deceived by people on social media. Therefore, this study contributes to the thoroughness of personality behavior based on their respective social media profile pictures. This project to study user behavior on social media profiles using CRIPS-DM methodology is suitable for using Random Forest and Decision Tree algorithm in Classification.

PENGENALAN

Di dalam era globalisasi teknologi yang semakin canggih, kewujudan media sosial seperti *Instagram*, *Facebook*, *Twitter*, *Tiktok* dan lain-lain melonjak naik kerana pengguna media sosial semakin meningkat tidak kira usia. Platform ini menyediakan dan memberi pengguna untuk berinteraksi, bertukar pendapat, mengenal seseorang daripada luar negara, berkongsi gambar, berkongsi gaya hidup dan sebagainya. Media sosial memberi tumpuan kepada orang dari seluruh dunia untuk bersosial bersama kenalan baru. Ini memudahkan penciptaan dan perkongsian kandungan yang dijana pengguna. Media sosial membantu pengguna membuat dan mencipta identiti tersendiri berdasarkan cara penyampaian pengguna media sosial. Ia juga boleh membangunkan tingkah laku pengguna dengan cara pengguna berkongsi gaya hidup dan kehidupan dengan orang lain. Cara tingkah laku pengguna menunjukkan diri mereka pada orang lain bergantung pada personaliti dan sejarah silam kehidupan masing-masing.

Perspektif psikologi dunia sangat popular dengan personaliti. Personaliti berkaitan dengan ciri penting tingkah laku pengguna dan secara relatifnya adalah tetap. Kajian gambar profil sangat menarik kerana ini adalah berkaitan dengan gambar yang dipilih oleh pengguna mewakili personaliti dalam talian mereka dan lebih-lebih lagi pengguna boleh menggunakan gambar bukan diri mereka. Pemilihan gambar profil di media sosial adalah pilihan pengguna dan dikaitkan dengan tingkah laku sebahagiannya daripada personaliti mereka.

Terdapat banyak platform media sosial dalam era globalisasi dan dunia teknologi yang berkongsi kehidupan sosial dan gaya hidup seseorang dengan hanya satu klik butang ikuti, walaupun mereka tidak diketahui. Terutamanya remaja, pelajar universiti yang berpengetahuan tinggi, mempunyai tahap rasa ingin tahu dan keinginan yang tinggi untuk meneroka perkara baharu. Akibatnya terdapat isu dengan akaun profil palsu dan tingkah laku yang tidak dijangka di media sosial.

KAJIAN LITERATUR

Personaliti seseorang boleh menjadi tarikan kepada orang lain untuk mengenal seseorang. Pada zaman yang berteknologi tinggi budaya mengenal seseorang melalui media sosial sering dilakukan dengan hanya melihat gambar disiarkan pada profil seseorang. Personaliti tingkah laku memberi impak dan peka untuk berkenalan dengan seseorang di media sosial. Sedia

maklum kelompok pengguna media sosial terutama *Instagram* adalah daripada kalangan remaja dan pelajar. Dengan pendedahan media sosial, dapat dilihat tingkah laku personaliti seseorang pada usia muda berdasarkan senarai mengikuti pada akaun *Instagram*. Dalam kajian ini, skop yang dipilih adalah tempat pengajian awam iaitu Universiti Pertahanan Nasional Malaysia kerana terdapat dua status pelajar iaitu kadet dan awam, seterusnya berlainan fakulti.

Berdasarkan data besar, artikel ini menggunakan algoritma pengumpulan terbahagi untuk mengkaji gelagat pengguna dalam talian (Wang, 2020). Artikel ini memilih pengguna dalam talian kolej sebagai objek penyelidikan dan kajian serta menganalisis tingkah laku dalam talian pengguna sekolah. Dengan mengkaji tingkah laku dalam talian pengguna, memahami kadar penggunaan sumber jalur lebar rangkaian kampus dan pengagihan penggunaan rangkaian, untuk menghalang pelajar daripada terlibat dalam dunia rangkaian maya, dan untuk memastikan pengguna rangkaian dapat meningkatkan pengalaman dalam talian rangkaian kampus semasa mengakses sumber rangkaian secara munasabah. Menurut Beynon (2018), pengumpulan ialah proses mengumpul statistik pada pelbagai objek abstrak atau realistik dan kemudian mengumpulkan objek yang serupa kepada pelbagai jenis. Secara ringkasnya, pengumpulan boleh membahagikan objek yang mempunyai persamaan ke dalam kumpulan yang sama dan kemudian mengelompokkan objek yang mempunyai persamaan.

Berdasarkan kajian Tadesse (2018) matlamat kajian ini adalah untuk meramal ciri personaliti pengguna *Instagram* berdasarkan ciri dan ukuran model *Big 5 Personality* yang berbeza. Teori *Big 5 Personality* tersebut dikaitkan dengan sifat keperibadian *OCEAN* yang memiliki akronim *Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism*. Meneliti kehadiran struktur rangkaian sosial dan ciri linguistik berbanding interaksi personaliti menggunakan set data projek *myPersonality*. Nowson (2006) mengklasifikasikan versi ekstra, kestabilan, kesesuaian dan ketelitian penulis blog menggunakan model ramalan *Naive Bayes* sebagai algoritma pembelajaran menggunakan set n-gram yang berbeza sebagai ciri. Menganalisis dan bandingkan empat model pembelajaran mesin dan lakukan korelasi antara setiap set ciri dan ciri personaliti. Menggunakan tiga algoritma pembelajaran mesin yang berbeza sebagai pendekatan dasar untuk model ramalan kepada data ciri yang berkaitan untuk perbandingan *Classification*. Menentukan algoritma dan melaksanakan dalam model ramalan untuk menerokai sejauh mana ia boleh meramalkan personaliti ciri

dari *Instagram* dan bandingkan algoritma yang tinggi ketepatan untuk *Big 5 Personality*.

Dalam kajian ini, Lie (2016) menganalisis bagaimana imej profil Twitter berbeza dengan personaliti pengguna menyiarkannya. Dalam analisis utama ia menggunakan imej profil daripada lebih 66,000 pengguna yang personalitinya kami anggarkan berdasarkan tweet mereka menggunakan algoritma *Linear Regression*. Untuk memudahkan tafsiran, ia menumpukan analisis pada estetika dan wajah ciri dan kawalan untuk variasi demografi dalam imej ciri dan personaliti. Extrovert lebih suka berinteraksi dengan orang ramai, dilihat bertenaga. Extrovert gemar menggunakan gambar profil yang mengandungi

orang lain atau memaparkan riak wajah yang lebih menyenangkan. Ketelitian pengguna sangat tinggi dan cenderung untuk lebih teratur dan lebih suka tingkah laku yang disiplin. Ini boleh menyebabkan pengguna mematuhi norma daripada apa yang diharapkan daripada profil gambar.

METODOLOGI

Kajian tingkah laku pengguna terhadap profil media sosial menggunakan metodologi *Cross Industry Standard Process for Data Mining (CRISP-DM)*.



RAJAH 1. Cross Industry Standard Process for Data Mining (CRISP-DM)

Cross Industry Standard Process for Data Mining (CRISP-DM) ialah model proses dengan enam fasa yang secara semula jadi menangkap kitaran hayat sains data (rujuk rajah 1). Ini pada asasnya satu set struktur untuk membantu, merancang, mengurus dan melaksanakan projek sains data atau pembelajaran mesin. Model ini mewakili urutan fasa yang ideal. Banyak aktiviti mungkin dijalankan dalam urutan yang berbeza dalam amalan dan tugas tertentu, ia juga sering disusuli dan diulang. Model ini tidak cuba menjejaki sebarang laluan semasa proses pengelasan data. Berikut adalah fasa-fasa didalam *CRISP-DM*:

- i. Pemahaman Kajian (*Research Understanding*).
- ii. Pemahaman Data (*Data Understanding*)
- iii. Penyediaan Data (*Data Preparation*)
- iv. Pemodelan (*Modelling*)
- v. Penilaian (*Evaluation*)
- vi. Pelaksanaan (*Deployment*)

Fasa pemahaman kajian memberi tumpuan kepada objektif dan keperluan projek. Matlamat peringkat proses ini adalah untuk mendedahkan faktor penting yang boleh mempengaruhi hasil projek.

Mengabaikan langkah ini bermakna banyak usaha yang perlu dilakukan untuk menghasilkan jawapan yang betul kepada soalan yang salah. Hasil daripada langkah ini, pengkaji pembangunan harus memahami konteks projek kajian. Di fasa ini untuk mengenal pasti keperluan data yang diperlukan untuk menganalisis tingkah laku pengguna. Memahami objektif kajian iaitu melakukan kajian berkaitan *introvert* dan *extrovert*.

Pada fasa pemahaman data, ia memberi pendedahan awal untuk memahami data tersebut sebelum melakukan proses penyediaan data. Pentingnya untuk memahami data yang diambil untuk mengkaji data tersebut. Jika pada fasa pemahaman data kurang memahami data, objektif kajian tidak dapat dilaksanakan dengan betul dan kurang tepat. Mengenal pasti data yang diambil bersesuaian dengan objektif kajian iaitu tentang personaliti *introvert* dan *extrovert* terhadap media sosial. Data diambil daripada pelajar berstatus awam dan kadet daripada Universiti Pertahanan Nasional Malaysia seramai 100 orang responden untuk melakukan kajian ini. Cara mengambil data tersebut melalui soalan-soalan tentang kekerapan pengguna menggunakan media sosial *Instagram* dan

objektif utama gambar profil media sosial pengguna. Borang soal selidik ini diedarkan melalui aplikasi *Whatsapp* pelajar di UPM. Borang soal selidik tersebut merangkumi 23 soalan yang dibahagikan kepada 4 bahagian iaitu bahagian data demografi, *Posting* di *Instagram*, komen di *Instagram* dan *Like* di *Instagram*.

Fasa penyediaan data merupakan fasa bagi melakukan proses pembersihan data sebelum melakukan proses pemodelan algoritma yang bersesuaian dengan data. Fasa ini amat penting kerana setiap data akan menjalani proses pembersihan data. Data yang memiliki masalah atau kesalahan akan dibersihkan supaya dapat tidak berlaku masalah Ketika melakukan analisis. Walau bagaimanapun, tidak bermakna pada fasa ini segala masalah pada data telah diselesaikan namun fasa ini akan digunakan untuk tinjauan sekiranya berlaku masalah pada data ketika pembangunan model.

Fasa pemodelan ini memainkan peranan untuk menentukan model yang akan digunakan sama ada meramal atau deskriptif. Fasa ini akan berjalan sekiranya data yang digunakan berada dalam keadaan baik. Jika perlu menyesuaikan data dengan suatu teknik perlombongan data dan algoritma maka, fasa penyediaan harus dilakukan semula. Antara model yang biasa digunakan adalah analisis *Classification* dan *Clustering*.

Bagi kajian ini, pada fasa pemodelan algoritma *Decision Tree* dan *Random Forest* digunakan. *Decision Tree* merupakan salah satu teknik *Classification* yang menggunakan struktur pokok di mana setiap nod mewakili atribut, di mana cabangnya mewakili nilai dari atribut, dan daun pula mewakili kelas. Nod yang paling atas dari *Decision Tree* disebut sebagai *root*. *Decision Tree* merupakan teknik *Classification* yang paling kerap digunakan. Selain pembangunan model yang cepat, hasil daripada model tersebut mudah untuk difahami. Selain itu algoritma pembelajaran mesin yang mudah untuk digunakan dan mampu menghasilkan keputusan yang baik adalah *Random Forest*. Algoritma tersebut juga sering digunakan kerana kesederhanaan dan variasi dalam menghasilkan keputusan bagi tugas klasifikasi dan regresi. *Random Forest* akan melakukan pemerhatian data secara rawak dan menghasilkan beberapa *Decision Tree* lalu mendapatkan purata bagi hasil tersebut. *Random Forest* memiliki suatu kelebihan iaitu hiperparameter yang dapat meningkatkan tahap peramalan model atau dapat menghasilkan model lebih pantas.

Pada fasa penilaian, hasil model yang dipilih akan dinilai dan diuji bagi memastikan keberkesannya untuk mencapai objektif kajian. Oleh itu, matlamat utama fasa ini adalah bagi memastikan kajian ini berada

dalam keadaan baik dan tiada masalah yang berlaku pada model tersebut. Walau bagaimanapun, sekiranya terdapat masalah dengan model maka masih boleh kembali ke fasa sebelumnya untuk menentukan model yang bersesuaian.

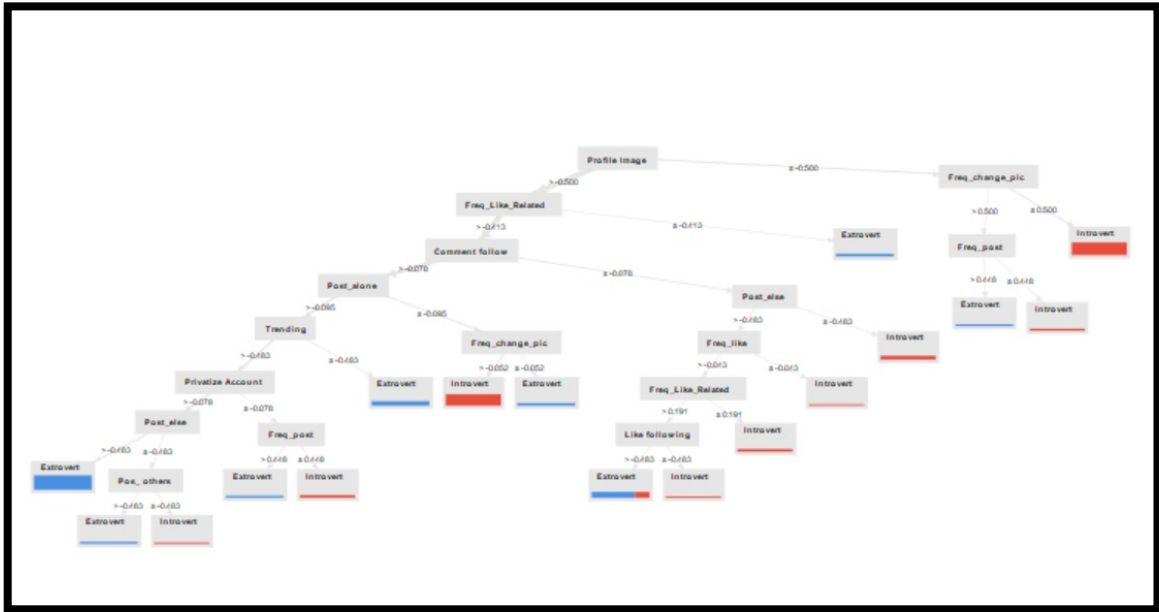
Bagi kajian ini, pada fasa penilaian juga akan dilakukan perbandingan antara algoritma *Decision Tree* dan *Random Forest* untuk memilih algoritma *Classification* yang terbaik. Perbandingan ini dilakukan dengan memilih algoritma yang memiliki ketepatan maksimum dan kekeliruan *Classification* minimum.

Fasa pelaksanaan merupakan fasa yang terakhir di mana hasil daripada fasa pemahaman kajian sehingga fasa penilaian akan dipaparkan. Hal ini kerana kesemua keperluan telah dipastikan dan proses analisis akan berjalan.

HASIL KAJIAN

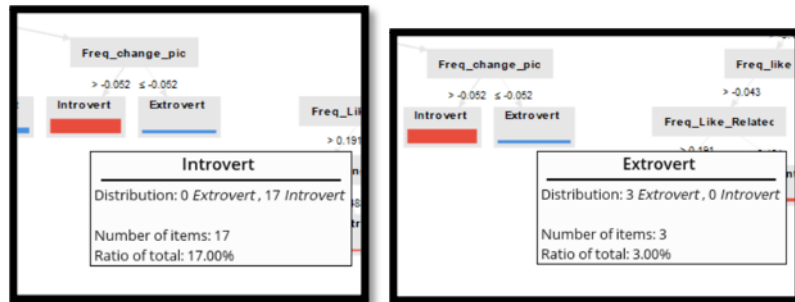
Algoritma *Random Forest* telah digunakan dalam proses pemodelan untuk menghasilkan keputusan kajian. Sebelum melakukan proses pemodelan, proses penilaian data dilakukan bagi mendapatkan skor setiap data dilakukan dengan menggunakan operator *Extract Sentiment*. Berdasarkan rajah 2, hasil analisa algoritma *Random Forest* menunjukkan bahawa imej profil telah ditetapkan sebagai tunjang utama bagi keseluruhan pokok kerana memiliki perolehan maklumat tertinggi. Hasil analisa ini merujuk kepada 20 soalan yang telah dipilih ketika proses penilaian data bagi pemilihan data yang terbaik bagi kajian ini. Setiap pokok individu dalam model *Random Forest* mengeluarkan ramalan kelas dan kelas dengan undian terbanyak menjadi model ramalan.

Rajah 2 menunjukkan hasil analisa algoritma *Random Forest* bagi menentukan pengguna *introvert* atau *extrovert*. Daripada nod *Profile Image* ia akan terbahagi kepada dua cabang iaitu *Freq_Like_related* dengan *Freq_change_pic*. Dapat dilihat nilai *gain ratio* daripada atribut *Profile Image* lebih besar daripada atribut *Freq_Like_related* > -0.500 dan atribut *Profile Image* nilai *gain ratio* lagi kurang daripada atribut *Freq_change_pic* ≤ -0.500 . Seterusnya nilai *gain ratio* lebih dari -0.413 untuk atribut *Freq_Like_related* manakala nilai *gain ratio* lebih dari -0.078 untuk atribut *Comment Follow* dan terbahagi kepada atribut *Post_else* dengan nilai *ratio* lebih kurang -0.078 . Untuk atribut *Post_alone* nilai *gain ratio* lebih dari -0.095 dan dibahagi dengan atribut *freq_change_pic* dengan nilai *gain ratio* lebih kurang -0.095 . Atribut *Freq_Change_pic* nilai *gain ratio* lebih dari -0.052 adalah *introvert*. Sebanyak 17 orang atau 17% adalah tingkah laku *introvert* dari



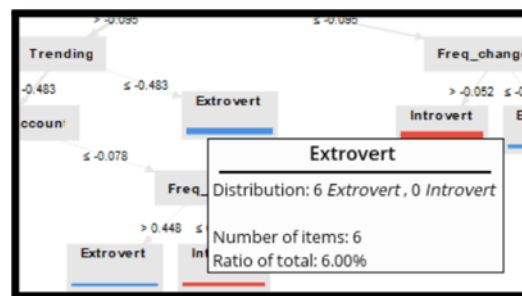
RAJAH 2. Hasil analisa algoritma Random Forest bagi menentukan pengguna introvert atau extrovert .

atribut kekerapan menukar profil gambar manakala nilai *gain ratio* lebih kurang dari -0.052 adalah dengan jumlah 3 orang atau 3% tingkah laku yang bersifat extrovert seperti pada rajah 3.



RAJAH 3. Analisa pelajar yang introvert dan extrovert .

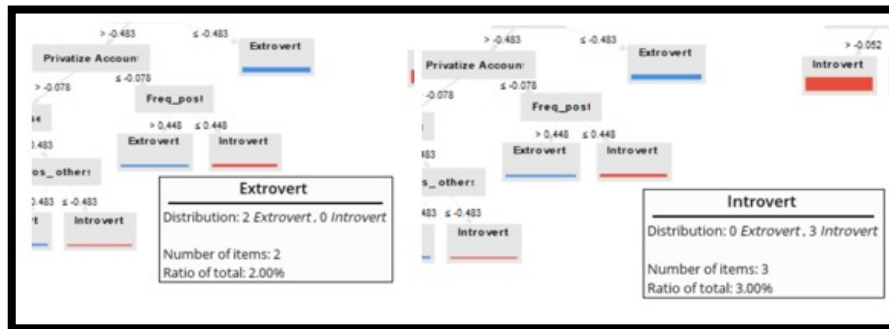
Berdasarkan rajah 4 nilai *gain ratio* untuk menganalisa tingkah laku *extrovert* yang sebanyak 6 atribut *Trending* lebih kurang -0.0483 dengan orang atau 6% .



RAJAH 4. Analisa pelajar extrovert .

Nilai *gain ratio* atribut *Privatize Account* lebih dari -0.078 dan di bawah kepada atribut *Freq_post* dengan nilai *ratio* lebih kurang -0.078. Atribut *Freq_post* menganalisa tingkah laku pengguna *extrovert* yang

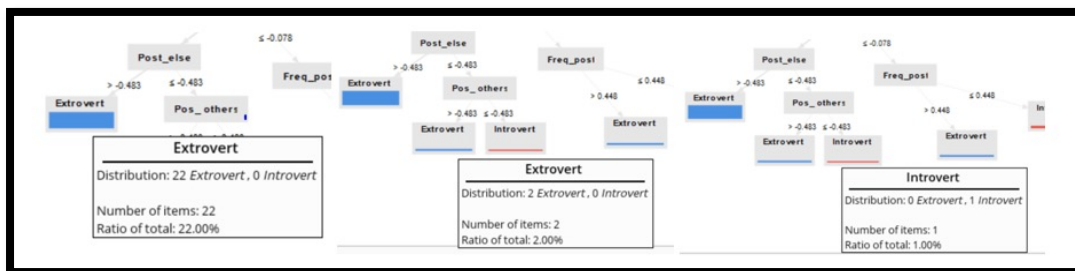
bernilai *gain ratio* lebih 0.048 dengan 2% manakala untuk tingkah laku pengguna *introvert* nilai *gain ratio* lebih kurang 0.048 dengan 3% seperti pada rajah 5



RAJAH 5. Analisa tingkah laku introvert dan extrovert di bawah atribut *Freq_post*.

Berdasarkan rajah 6, menunjukkan untuk nilai *gain ratio* atribut *Post_else* lebih dari -0.0483 dan menganalisa bahawa 22% adalah tingkah laku *extrovert* daripada atribut *Post_else*. Atribut *Post_others* membawa nilai *gain ratio* sama seperti *Post_*

Else iaitu lebih -0.0483 dan menganalisa sebagai *extrovert* dengan 2% manakala nilai *gain ratio* untuk menganalisa introvert adalah lebih kurang -0.0483 dengan 1%.



RAJAH 6. Analisa tingkah laku di bawah atribut *Post_Else* dan *Post_Others*.

Rajah 7 menunjukkan bahagian kanan pokok yang menggunakan *attribute* nod yang sama iaitu *Profile Image*. Nilai *gain ratio* untuk atribut *Freq_like_related post* lebih kurang dari -0.0413 yang membawa kepada 3% dan ini menunjukkan analisa kekerapan pengguna untuk *like posting* dengan keadaan emosi yang berkenan dengan diri mereka adalah 3% membawa kepada tingkah laku yang *extrovert* seperti rajah 8.

Atribut *Post_else* memegang nilai *gain ratio* lebih kurang -0.0483 dan analisa menunjukkan sebanyak 5% adalah tingkah laku pengguna yang introvert. Ini menunjukkan 5% ini tidak gemar untuk memperlihatkan orang yang mereka rapat dan sayang diperlihatkan oleh rakan-rakan di media sosial. Seperti atribut *Freq_like* nilai *gain ratio* lebih kurang

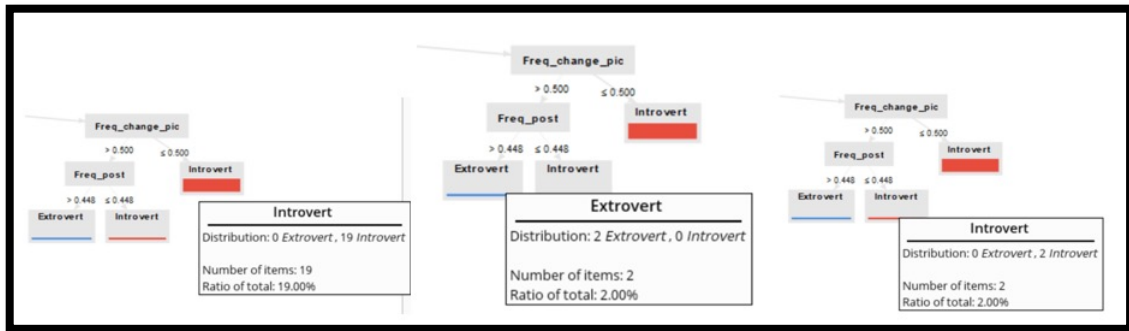
-0.043 yang membawa kepada 1% sahaja untuk tingkah laku *introvert*. Ini menunjukkan seseorang itu benar membawa perwatakan ciri-ciri *introvert* yang tidak gemar untuk melihat media sosial *Instagram*, selebihnya membawa kepada ciri-ciri *extrovert* yang like posting rakan-rakan mereka di *Instagram*. Rajah 9 menunjukkan *subtree* atribut *Post_else* dan *Freq_like*.

Pada rajah 10, di atribut *Freq_Like_related* nilai *gain ratio* lebih dari 0.191 daripada atribut *Like Following* manakala *gain ratio* lebih kurang 0.191 menganalisis tingkah laku pengguna yang *introvert* membawa kepada 3%. 3% ini tingkah laku yang tidak gemar untuk *like posting* yang mempunyai persamaan keadaan semasa pengguna. Atribut *Like_Following* mendapat *gain ratio* lebih -0.0483 yang membawa

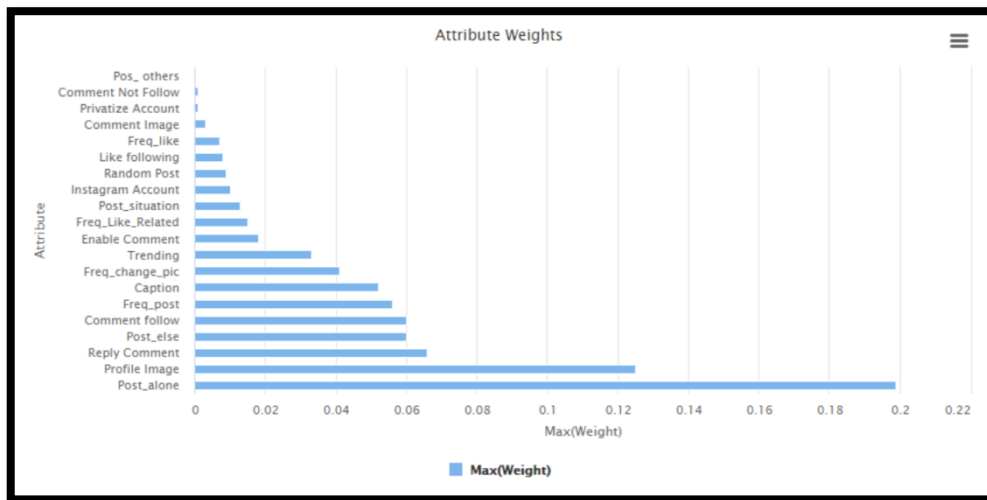
kepada tingkah laku yang *extrovert* dengan 8% daripada mereka suka *like posting* rakan-rakan tetap di *Instagram* manakala untuk *introvert* membawa kepada *gain ratio* lebih kurang -0.0483 bersamaan 2% yang tidak gemar *like posting* rakan-rakan di *Instagram*.

Akhir sekali untuk atribut *Freq_Change_Picture* seperti rajah 11 mendapat *gain ratio* lebih kurang 0.500 untuk tingkah laku *introvert*. Ini

membawa kepada 19% daripada responden tidak gemar untuk menukar profil gambar di *Instagram*. Manakala untuk atribut *Freq_Post* *gain ratio* lebih 0.048 tingkah laku *extrovert* yang 2% gemar untuk post gambar setiap minggu mahupun setiap bulan manakala untuk *introvert* yang mendapat *gain ratio* lebih kurang 0.048, 2% tidak gemar *post* gambar di *Instagram*.



RAJAH 11. Analisa untuk atribut *Freq_Change_Pic*.



RAJAH 12. Atribut *weight* untuk setiap soalan.

Berdasarkan rajah 12, *Attribute Weight* digunakan bagi menentukan nod bagi mewujudkan model *Random Forest*. Atribut yang memiliki berat yang tinggi akan menjadi nod pokok bagi model *Random Forest* manakala yang rendah akan berada pada bahagian bawah pokok tersebut. Secara keseluruhan, analisa hasil kajian ini adalah nilai skor daripada *Random Forest* menunjukkan tingkah laku *introvert* melebihi *extrovert*.

Keputusan ketepatan hasil kajian tingkah

laku pengguna terhadap 100 responden akan dilihat dan membandingkan ketepatan dengan menggunakan algoritma yang berbeza. Ketepatan dikira dengan mengambil peratusan ramalan yang betul berbanding jumlah bilangan data tingkah laku. Berdasarkan rajah 13, ketepatan untuk hasil kajian tingkah laku adalah 79%. Di kelas sebenar untuk ramalan *extrovert* mempunyai *true extrovert* yang merujuk kepada *tupel positif (extrovert)* yang telah dilabelkan dengan betul oleh pengelas adalah 33 manakala untuk *true introvert*

yang merujuk kepada *tupel negative (introvert)* yang telah dilabelkan dengan betul oleh pengelas adalah 46. Seterusnya untuk kelas sebenar ramalan *introvert* mempunyai nilai *false negative* di *true extrovert* yang merujuk *tupel* positif yang telah disalah labelkan

sebagai negatif oleh pengelas adalah 11 dan untuk *false positive* merujuk di *true introvert* adalah *tupel* negatif yang telah dilabel secara salah sebagai positif dengan nilai 10.

```

PerformanceVector:
accuracy: 79.00% +/- 11.01% (micro average: 79.00%)
ConfusionMatrix:
True:  Extrovert  Introvert
Extrovert :    33     10
Introvert :    11     46
classification_error: 21.00% +/- 11.01% (micro average: 21.00%)
ConfusionMatrix:
True:  Extrovert  Introvert
Extrovert :    33     10
Introvert :    11     46

```

RAJAH 13. Ketepatan Paparan *PerformanceVector*.

JADUAL 1. Confusion Matrix.

Predicted Value	Actual Value	
	Extrovert	Introvert
Extrovert	33 (TP)	10 (FP)
Introvert	11 (FN)	46 (TN)

Berdasarkan rajah 13 extrovert bersamaan dengan *True Positive* (TP) dan introvert bersamaan dengan *True Negative* (TN). Manakala positif bersamaan dengan *extrovert* dan negatif bersamaan dengan introvert.

Berdasarkan Jadual 2, ketepatan class yang tinggi adalah ramalan *introvert* yang nilainya 80.70%

dan ketepatan kelas yang rendah adalah ramalan *extrovert* dengan nilai 76.74%. Untuk *class recall true introvert* yang nilainya 82.14 % lagi tinggi daripada *true extrovert* yang nilai 75%. Berdasarkan keputusan hasil kajian ini, pelajar mempunyai tingkah laku yang *introvert* lagi tinggi daripada tingkah laku *extrovert*.

Hasil kajian yang telah dilakukan terdapat

JADUAL 2. Pengiraan berdasarkan Confusion Matrix.

Kriteria	Formula	Pengiraan	Peratusan
Accuracy	$TP+TN / (TP+TN+FP+FN)$	$33 + 46 / (33 + 10 + 11 + 46) = 0.79$	79%
Sensitivity	$TP / (TP+FN)$	$33 / 33 + 11 = 0.75$	75%
Specificity	$TN / (TN+FP)$	$46 / 46 + 10 = 0.8214$	82.14%
Precision	$TP / (TP+FP)$	$33 / 33 + 10 = 0.7674$	76.74%
Recall	$TP / (TP+FN)$	$33 / 33 + 11 = 0.75$	75%

perbezaan tingkah laku dan prestasi diri pengguna media sosial *Instagram* yang memiliki keperibadian *extrovert* dan *introvert*. Hasil tersebut diketahui bahawa pengguna media sosial *Instagram* yang tingkah laku introvert memiliki rata-rata lebih tinggi pada semua

prestasi diri mereka dibandingkan dengan pengguna media sosial *Instagram* yang bertingkah laku *extrovert*. Tambahan pula, hasil analisa pada perisian *Rapid Miner* menunjukkan data mengenai *Caption* untuk *posting*, personaliti dan tingkah laku introvert didapati

lebih tinggi berbanding tingkah laku *extrovert*. Hal ini menjelaskan bahawa jenis personaliti *introvert* lebih gemar untuk berkomunikasi di media sosial berbanding bersemuka (Voorn & Kommer, 2013). Selain itu, jenis personaliti tingkah laku yang mempunyai sifat pemalu, kurang berkomunikasi dengan kawan seperti membalas komen di *Instagram*, menggunakan strategi penyampaian sendiri iaitu dengan meletak *caption* pada *post Instagram*.

Selain itu, pengguna *Instagram* personaliti dan tingkah laku *extrovert* memberi kesan kepada orang lain memahami diri mereka sebagai berkebolehan, berkemahiran dan berkelayakan. Selain itu, tingkah laku *extrovert* juga memiliki sifat optimis dan lebih bersemangat berbanding jenis personaliti *introvert* (Eysench & Suryabrata, 2003). Menurut McCabe KO & Fleeson W (2012), *extrovert* lebih cenderung untuk menegaskan diri mereka dalam kumpulan, mereka sering berperanan sebagai pemimpin apabila bekerja dengan orang lain. Ini juga sinonim dengan RQ7 iaitu data “*Post picture with someone else*” personaliti dan tingkah laku *extrovert* sering menunjukkan diri mereka selesa bersama dengan orang ramai dan nilai kebarangkalian $P(A)$ *extrovert* lebih tinggi daripada personaliti tingkah laku *introvert*.

Personaliti dan tingkah laku *extrovert* memperoleh tenaga dengan terlibat dalam interaksi sosial. Tingkah laku *extrovert* yang tinggi memerlukan rangsangan sosial untuk berasa bertenaga. Personaliti

ini mendapat inspirasi dan keseronokan untuk bertegur dan bercakap dengan rakan-rakan media sosial. Berkait dengan RQ ‘*comment with following*’ dan mendapat nilai kebarangkalian 0.0091 lebih tinggi daripada personaliti tingkah laku *introvert*. Personaliti dan tingkah laku *introvert* lebih suka untuk menyiarkan gambar seorang diri berbanding dengan dengan orang lain kerana mereka berasa tidak selesa untuk mempertontonkan orang yang mereka selesa di *Instagram* dan ini bersesuaian dengan RQ ‘*post alone*’ yang menunjukkan nilai std *introvert* 0.0949 lebih tinggi daripada *extrovert* std 0.061.

PERBINCANGAN

Matlamat utama perbandingan ini adalah untuk mencari algoritma yang mempunyai tahap prestasi yang tinggi dan tahap kekeliruan *Classification* yang rendah. *Random Forest* dan *Decision Tree* ialah dua teknik perlombongan data yang kerap digunakan kerana ia memberikan hasil yang terbaik. *Random Forest* mempunyai ketepatan yang lebih tinggi daripada *Decision Tree*, mengikut perbandingan ini dengan ketepatan 79% dimiliki oleh *Random Forest* manakala untuk kekeliruan *Classification* juga rendah iaitu 21% berbanding *Decision Tree*. Berikut merupakan perbandingan penggunaan algoritma *Random Forest* dan *Decision Tree* untuk kajian ini.

JADUAL 3. Perbandingan Algoritma.

Algoritma	Random Forest	Decision Tree
Ketepatan (accuracy)	79%	76%
Kekeliruan Klasifikasi	21%	24%
Sensitivity	75%	79%
Specificity	82.14%	73.21%
Precision	76.74%	70%
Recall	75%	79.55%
Pemahaman Model	Lebih banyak pilihan	Pilihan yang tetap

Algoritma *Random Forest* mempunyai pilihan *tree* yang banyak berdasarkan *number of folds* pemilihan hasil model yang terbaik dapat dilakukan. Disebabkan *Random Forest* mengeluarkan *number of folds* yang banyak, ketetapan hasil kajian lebih tinggi daripada *Decision Tree*. Manakala untuk *Decision Tree* hanya memilih atribut yang terbaik iaitu *Post_alone* tetapi *Post_alone* tidak menepati hasil utama kajian.

Hasil utama kajian adalah untuk mengkaji *Profile Image* sama ada tingkah laku *introvert* atau *extrovert*. Berdasarkan pemahaman model di jadual 3, pemahaman model untuk *Random Forest* lebih banyak pilihan, ini bermakna dengan pengeluaran *tree* atau *number of folds* yang banyak dan untuk *Decision Tree* hanya membuat pilihan yang tetap. Disebabkan *number of folds* atau *tree* yang banyak di *Random Forest*, hasil tersebut iaitu

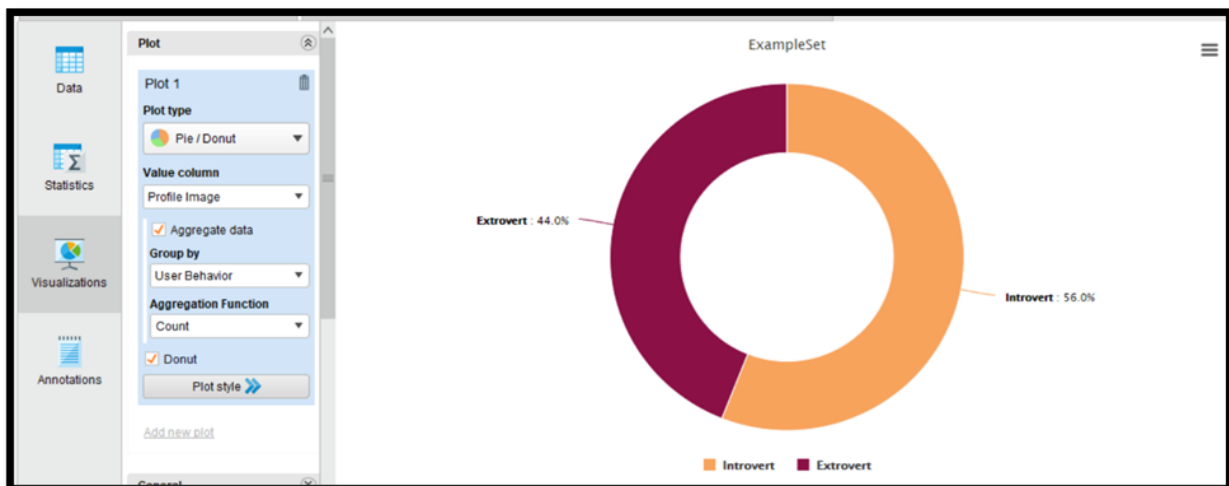
tree boleh memilih berdasarkan atribut dan kajian yang dipilih. Kaedah *Random Forest* mengeluarkan banyak pilihan pokok untuk *Classification* daripada atribut dengan undian dan pengiraan yang majoriti (Oshiro, 2012).

Seterusnya berdasarkan jadual 3, salah satu faktor *introvert* lebih tinggi daripada *extrovert* tingkah adalah *introvert* gemar melakukan aktiviti dan penggunaan di *Instagram* kerana tingkah laku yang *introvert* tidak gemar bersosial secara bersemuka

(Voorn & Kommer, 2013). Berdasarkan hasil kajian, berkemungkinan personaliti *introvert* akan bertambah kerana penggunaan media sosial tidak perlu bersemuka hanya berinteraksi melalui telefon pintar atau gajet. Salah satu objektif kajian ini adalah untuk mengenal pasti *posting* dan aktiviti penggunaan di *Instagram* bagi menggambarkan ciri-ciri personaliti pengguna sama ada *introvert* atau *extrovert* berdasarkan borang soal selidik bagi mendapatkan data.

JADUAL 4. Perbandingan algoritma melibatkan ketepatan

Algoritma	Random Forest	Decision Tree
User Behavior	Extrovert Introvert 33 46	Extrovert Introvert 35 41
Accuracy	79%	76%



RAJAH 14. Carta Pi antara *Profile Image* dengan *User Behavior*

Jadual 4, menunjukkan ketepatan kelas terhadap tingkah laku berserta algoritma untuk memudahkan kefahaman kajian ini. Bagi algoritma *Random Forest*, *extrovert* mempunyai nilai 33 lebih rendah berbanding *introvert* yang mempunyai nilai 46 berdasarkan daripada data borang soal selidik dan pengiraan *score*. Seterusnya bagi algoritma *Decision Tree* nilai *extrovert* 35 manakala *introvert* 41 berdasarkan pengiraan *score* dan data borang soal selidik. Hasil keputusan ketepatan nilai tingkah laku tersebut sebanyak 79%. Ini menunjukkan hasil keputusan tersebut tinggi dan memuaskan. Pada rajah 14, menunjukkan *Profile Image* untuk *extrovert* 44% dan *introvert* 56% dengan

di bawah penggunaan algoritma *Random Forest*. Manakala algoritma *Decision Tree* memegang nilai ketepatan 76% iaitu lebih rendah daripada *Random Forest*.

RUMUSAN

Pencapaian dalam membangunkan kajian ini boleh diukur melalui objektif yang digariskan. Objektif yang dicapai dalam kajian tingkah laku pengguna terhadap profil media sosial (*Instagram*) adalah untuk menyiasat gambar profil pengguna di sebalik akaun media sosial

mereka dan ciri-ciri pengguna berdasarkan gambar profil yang digunakan. Mengenal pasti posting di *Instagram* menggambarkan ciri-ciri personaliti pengguna sama ada introvert atau *extrovert*. Menganalisis dan meramal gambar profil pengguna dalam media sosial boleh menggambarkan tingkah laku mereka sama ada introvert atau *extrovert*. Antara kelebihan hasil kajian tingkah laku pengguna terhadap profil media sosial (*Instagram*) adalah dapat membezakan personaliti tingkah laku pengguna berdasarkan akaun media sosial *Instagram* menerusi komen, gambar, posting dan like. Kajian ini juga dapat mengenal lebih jelas jenis personaliti pengguna daripada hasil kajian dan hasil kajian literatur.

RUJUKAN

- Adamopoulos, P., Ghose, A., & Todri, V. (2018). The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms. *Information Systems Research*, 29(3), 612–640. <https://doi.org/10.1287/isre.2017.0768>
- Amichai-Hamburger, Y., Wainapel, G., & Fox, S. (2002). “On the Internet No One Knows I’m an Introvert”: Extroversion, Neuroticism, and Internet Interaction. *CyberPsychology & Behavior*, 5(2), 125–128. <https://doi.org/10.1089/109493102753770507>
- Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. (2002). Can You See the Real Me? Activation and Expression of the “True Self” on the Internet. *Journal of Social Issues*, 58(1), 33–48. <https://doi.org/10.1111/1540-4560.00247>
- Chen, Y., Liu, C., Du, Y., Zhang, J., Yu, J., & Xu, H. (2022). Machine learning classification model using Weibo users’ social appearance anxiety, Personality and Individual Differences, 188, 111449. <https://doi.org/10.1016/j.paid.2021.111449>
- Conklin, T. (2014). *Social and Emotional Learning in Middle School: Essential Lessons for Student Success: Engaging Lessons, Strategies, and Tips That Help Students . . . Navigate Middle School and Focus on Academics*. Scholastic Teaching Resources (Teaching).
- Liu, L., Preotiuc-Pietro, D., Riahi Samani, Z., E. Moghaddam, M., & Ungar, L. (2021). Analyzing Personality through Social Media Profile Picture Choice. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), 211–220. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/14738>
- Lukito, L. C., Erwin, A., Purnama, J., & Danoekoesoemo, W. (2016). Social media user personality classification using computational linguistic. *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*. Published. <https://doi.org/10.1109/iciteed.2016.786331>
- McCay-Peet, L., & Quan-Haase, A. (2016). A Model of Social Media Engagement: User Profiles, Gratifications, and Experiences. *Why Engagement Matters*, 199–217. https://doi.org/10.1007/978-3-319-27446-1_9
- Mushtaq, Z., Ashraf, S., & Sabahat, N. (2020). Predicting MBTI Personality type with K-means Clustering and Gradient Boosting. *2020 IEEE 23rd International Multitopic Conference (INMIC)*. <https://doi.org/10.1109/inmic50486.2020.9318078>
- Rarasati, D. B., Hudaniah, H., & Prasetyaningrum, S. (2019). Strategi presentasi diri pengguna instagram ditinjau dari tipe kepribadian. *Jurnal Ilmiah Psikologi Terapan*, 7(2). <https://doi.org/10.22219/jipt.v7i2.8164>
- Setiawan, H., & Wafi, A. A. (2020). Classification of Personality Type Based on Twitter Data Using Machine Learning Techniques. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*. <https://doi.org/10.1109/icoiaact50329.2020.9332152>
- Sheeks, M. S., & Birchmeier, Z. P. (2007). Shyness, Sociability, and the Use of Computer-Mediated Communication in Relationship Development. *CyberPsychology & Behavior*, 10(1), 64–70. <https://doi.org/10.1089/cpb.2006.9991>
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality Predictions Based on User Behavior on the Facebook Social Media Platform. *IEEE Access*, 6, 61959–61969. <https://doi.org/10.1109/access.2018.2876502>
- Wang, Y. (2020). User online behavior based on big data distributed clustering algorithm. *International Journal of Advanced Robotic Systems*, 17(2), 172988142091729. <https://doi.org/10.1177/1729881420917293>
- Weinschenk, A. C. (2017). Big Five Personality Traits, Political Participation, and Civic Engagement: Evidence from 24 Countries. *Social Science Quarterly*, 98(5), 1406–1421. <https://doi.org/10.1111/ssqu.12380>

Siti Ubaidah Ramlan, Suzaimah Ramli* & Shahaneim Marzukhi
Jabatan Sains Komputer, Fakulti Sains dan Teknologi Pertahanan,
Universiti Pertahanan Nasional Malaysia,
Kem Sg Besi 57000 Kuala Lumpur.

*Pengarang untuk surat menyurat; e-mel: suzaimah@upnm.edu.my

Diserahkan: 14 April 2022

Diterima: 11 Mei 2022