

PERLOMBONGAN BERITA BERKAITAN JENAYAH SIBER DARI WEB

Nur Alyya Izzati binti Shaharul Azhar

Dr. Wan Fariza Paizi@Fauzi

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Peningkatan jenayah siber adalah seiring dengan kemajuan rangkaian manusia dan mesin. Jenayah siber adalah sebarang aktiviti jenayah yang dilakukan menggunakan teknologi komputer menerusi rangkaian internet. Kes jenayah siber amatlah memberi kesan kepada mangsa dan bukan disebabkan masalah kewangan sahaja. Laman web mengenai jenayah siber digunakan secara meluas di luar negara. Oleh sebab Malaysia masih belum mempunyai laman web yang memaparkan mengenai kes jenayah siber, sebuah laman web akan dibangunkan bagi menghimpun segala berita dan maklumat mengenai jenayah siber bagi membolehkan situasi dan peristiwa semasa diketahui umum. Objektif kajian ini adalah untuk memberi kesedaran kepada masyarakat Malaysia tentang bahayanya jenayah siber. Berita jenayah siber akan dilombong secara automatik daripada akhbar online tempatan merangkumi akhbar online dalam Bahasa Malaysia. Kata kunci jenayah siber diperlukan untuk mendapatkan berita yang tepat. Selain itu, senarai laman akhbar tempatan telah ditetapkan agar proses perlombongan berita dapat berjalan dengan lancar. Seterusnya berita tersebut akan dikumpul mengikut kategori menggunakan kaedah 'k-means klustering'. Sistem ini dibangunkan menggunakan metodologi Model Kitar Hayat Pembangunan Sistem. Hasil laman web yang dibangunkan adalah untuk menyelesaikan masalah yang dihadapi serta mencapai objektif yang dinyatakan.

1 PENGENALAN

Jenayah siber seringkali berlaku pada masa kini sejajar dengan pembangunan teknologi yang semakin canggih. Jenayah didefinisi sebagai perbuatan jahat seperti mencuri, merompak dan membunuh yang salah di sisi undang-undang (Kamus Dewan Edisi Keempat). Perkataan siber pula didefinisi sebagai perbuatan atau hal yang berkaitan dengan komputer atau Internet (Kamus Pelajar Edisi Kedua).

Jenayah siber mempunyai dua kategori iaitu jenayah yang berlaku terhadap komputer atau peranti lain seperti penggodaman. Jenayah siber juga merupakan jenayah di mana komputer atau peranti lain adalah menjadi subjek untuk melakukan sesuatu kejahatan seperti

penipuan dalam talian, kecurian identiti dan pengedaran bahan eksploitasi kanak-kanak (Acorn 2018).

Portal jenayah siber merupakan sebuah laman web yang mengumpulkan segala berita atau maklumat berkaitan jenayah siber. Portal ini digunakan secara meluas di luar negara di mana setiap negara mempunyai satu portal jenayah siber untuk mengumpulkan segala berita jenayah siber yang berlaku di negara tersebut. Kajian ini membincangkan perlombongan berita berkaitan jenayah siber dari web.

2 PENYATAAN MASALAH

Jenayah siber merupakan suatu jenayah yang membahayakan. Hal ini kerana jenayah siber boleh berlaku kepada sesiapa sahaja tidak mengenal umur dan pangkat. Selalunya jenayah siber berlaku melibatkan wang tunai dan nyawa seseorang. Walaupun berita jenayah siber ini terpapar di akhbar tempatan online mahupun surat khabar, rakyat Malaysia seakan-akan tidak mengambil iktibar dan berjaga-jaga dengan kes jenayah siber yang berlaku. Rakyat Malaysia masih tidak berhati-hati sewaktu memberi maklumat akaun peribadi kepada seseorang yang dikenali. Contohnya, jika satu nombor yang berlakon menjadi anggota polis ataupun bank menelefon mereka untuk meminta nombor kad pengenalan, mereka sewenang-wenangnya memberi maklumat itu tanpa membuat siasatan terlebih dahulu. Oleh yang demikian, mereka telah menjadi mangsa jenayah siber yang dikenali sebagai penipuan ataupun scam.

Perlombongan berita berkaitan jenayah siber dari web merupakan salah satu langkah alternatif untuk mengurangkan kadar kes jenayah siber. Satu laman web akan dibangunkan bagi mengumpul segala berita berkaitan jenayah siber yang berlaku di Malaysia.

3 OBJEKTIF KAJIAN

Projek ini bertujuan untuk mendedahkan kes jenayah siber yang berlaku di Malaysia. Masyarakat perlu tahu mengenai jenayah siber sebagai langkah berjaga-jaga daripada menjadi mangsa jenayah siber.

Secara umum, objektif kajian ini adalah untuk mengumpul segala berita dan maklumat mengenai jenayah siber yang berlaku di Malaysia secara berkala dan automatik. Kajian ini juga

memaparkan statistik kes jenayah siber supaya masyarakat Malaysia menjadi lebih peka dan berhati-hati akan jenayah siber yang sering berlaku di Malaysia.

4 METOD KAJIAN

Penggunaan model pembangunan yang sesuai penting untuk memastikan perjalanan projek dapat berjalan dengan lancar dan menjamin hasil kerja yang berkualiti. Model perlombongan berita berkaitan jenayah siber dari web melibatkan beberapa fasa pembangunan serta penggunaan perisian dan perkakasan yang bersesuaian. Fasa pembangunan termasuk fasa perancangan sistem, analisis, reka bentuk, pelaksanaan serta sokongan dan keselamatan sistem.

4.1 Fasa Perancangan Sistem

Fasa ini merupakan fasa di mana proses untuk membangunkan sistem akan dirancang mengikut masalah yang dihadapi. Perancangan sistem amatlah penting supaya sistem yang dibangunkan dapat bertahan lama dan bukan sekadar sebuah sistem yang sia-sia.

4.2 Fasa Analisis Sistem

Fasa analisis data merupakan fasa di mana kita menganalisis segala keperluan perisian dan maklumat yang diperolehi dalam membangunkan sistem yang berguna. Analisis data penting untuk mengetahui cara yang betul untuk memproses data mengikut data yang diperolehi.

4.3 Fasa Reka Bentuk Sistem

Fasa ini merupakan fasa di mana idea reka bentuk bagi sistem yang akan dibangunkan telah ada. Sistem direka bentuk mengikut kesesuaian skop kajian di mana sistem ini akan diakses oleh pengguna dari pelbagai latar belakang. Oleh itu, reka bentuk sistem yang dibangunkan mestilah mesra pengguna dan senang faham. Bagi menghasilkan reka bentuk sistem ini, Bootstrap telah digunakan bagi menghasilkan reka bentuk sistem yang mesra pengguna dan senang digunakan.

4.4 Fasa Pelaksanaan Sistem

Di fasa ini, sistem yang telah mempunyai data dan reka bentuk akan dilaksanakan kepada pengguna. Sistem ini akan diuji untuk mengetahui samada reka bentuk dan fungsi yang telah dibangunkan sesuai dan senang digunakan oleh pengguna. Kelancaran sistem juga amatlah penting bagi sesebuah sistem yang akan dibangunkan.

Perkakasan dan perisian yang digunakan untuk menghasilkan projek haruslah dipilih dengan teliti. Perisian dan perkakasan yang baik dapat berfungsi dengan baik dan lancar serta menyokong perlombongan berita berkaitan jenayah siber dari web. Pemilihan perkakasan dan perisian yang salah akan mengakibatkan kegagalan dalam membangunkan projek ini. Spesifikasi perkakasan yang digunakan adalah perkakasan asas sebuah komputer. Senarai spesifikasi perkakasan yang dicadangkan untuk membangunkan projek ini adalah seperti berikut:

- i. Sistem Pengoperasian: Microsoft Windows 10
- ii. Pemprosesan: Intel® Core™ i5-7200U CPU @ 2.50GHz 2.71 GHz
- iii. Ruang Cakera Keras (*Hardisk*): 1TB
- iv. Ingatan Cakera Rawak (*RAM*): 4GB
- v. Kad Grafik: NVIDIA GeForce 930MX, with GDDR5, 2GB VRAM

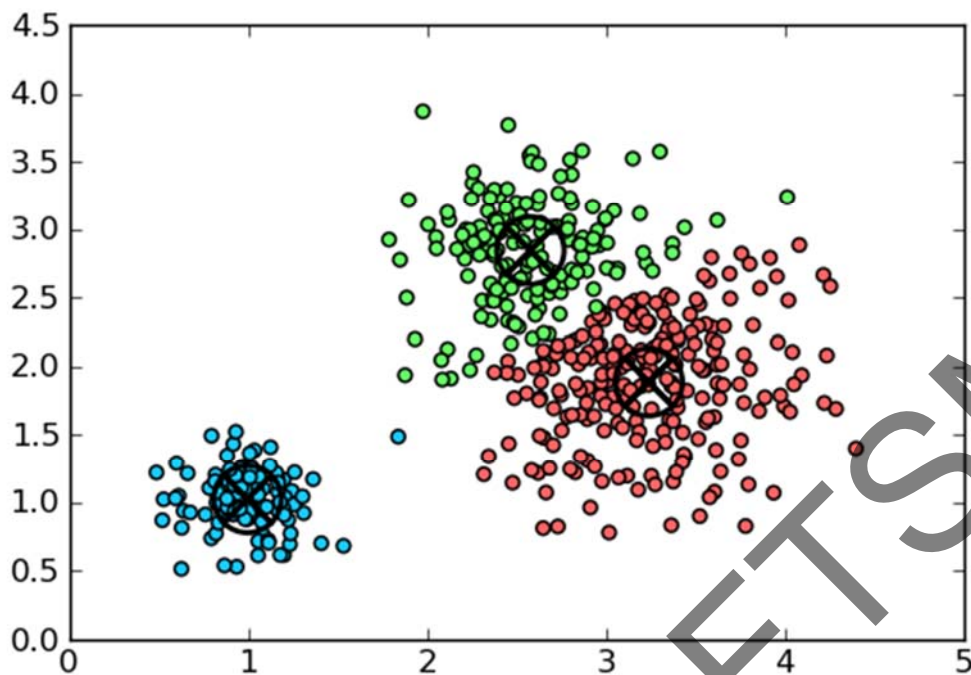
Spesifikasi perisian yang digunakan bagi membangunkan projek ini adalah Bootstrap, Scrapy dan bahasa pengaturcaraan Python serta k-means clustering. Bootstrap merupakan sebuah toolkit daripada sumber terbuka dengan rangka kerja HTML, CSS, dan JS yang paling popular untuk membangunkan projek pertama mudah alih yang responsif di web.

Seterusnya, Scrapy merupakan rangka kerja daripada sumber terbuka dan mempunyai kolaborasi untuk mengekstrak data yang anda perlukan dari laman web. Scrapy juga sebuah perisian yang cepat, ringkas dan mudah. Bahasa pengaturcaraan yang digunakan untuk membangunkan sistem ini ialah python. Python merupakan bahasa pengaturcaraan daripada sumber terbuka.

K-means clustering adalah sejenis pembelajaran tanpa pengawasan, yang digunakan apabila anda mempunyai data tidak berlabel (iaitu, data tanpa kategori atau kumpulan yang ditentukan). Matlamat algoritma ini adalah untuk mencari kumpulan dalam data, dengan bilangan kumpulan yang diwakili oleh pembolehubah K. Algoritma berfungsi secara berulang untuk memberikan setiap titik data kepada salah satu kumpulan K berdasarkan ciri-ciri yang disediakan. Titik data dikumpulkan berdasarkan persamaan ciri.

Hasil algoritma kluster K-means adalah:

- i. Centroids cluster K, yang boleh digunakan untuk melabelkan data baru
- ii. Label untuk data latihan (setiap titik data ditugaskan ke satu kelompok)



Rajah 1 Contoh pengumpulan data menggunakan k-means clustering

4.5 Fasa Sokongan dan Keselamatan Sistem

Fasa ini merupakan fasa terakhir dalam Model Kitar Hayat Pembangunan Sistem. Pada fasa ini, segala masalah yang berlaku semasa pelaksanaan sistem akan diperbetul mengikut kesesuaian cadangan yang dinyatakan oleh pengguna sewaktu sistem itu diuji buat kali pertama.

5 HASIL KAJIAN

Bahagian ini membincangkan hasil daripada proses perlombongan berita berkaitan jenayah siber dari web. Sistem ini dibangunkan secara berperingkat. Pada peringkat awal, berita berkaitan jenayah siber akan dilombong daripada akhbar tempatan online. Kata kunci 'siber' telah digunakan untuk mendapatkan berita yang betul dan tepat.

The screenshot shows the Astro Awani website interface. At the top, there are navigation links for 'BERITA', 'VIDEO', 'FOTO', and 'LIVE'. Below the navigation, there is a search bar and a 'CARI' button. The main content area displays a grid of news articles. Each article includes a thumbnail image, a headline, and a date. On the right side, there is a sidebar with two sections: 'TERKINI' (Latest) and 'POPULAR' (Popular). The 'TERKINI' section lists six news items with their respective categories (DUNIA, MALAYSIA, SUKAN) and headlines. The 'POPULAR' section is currently empty.

Rajah 2 Contoh akhbar tempatan online

```

C:\Users\User\Desktop\FinalYearProject\awaninews>scrapy crawl awani -o news.csv
2018-05-04 23:30:06 [scrapy.utils.log] INFO: Scrapy 1.5.0 started (bot: awaninews)
2018-05-04 23:30:06 [scrapy.utils.log] INFO: Versions: lxml 4.1.0.0, libxml2 2.9.4, cssselect 1.0.3, parsel 1.4.0, w3lib 1.19.0, Twisted 17.9.0, Python 2.7.14 |Anaconda
custom (64-bit) (default, Oct 15 2017, 03:34:40) [MSC v.1500 64 bit (AMD64)], pyOpenSSL 17.2.0 (OpenSSL 1.0.2l 25 May 2017), cryptography 2.0.3, Platform Windows-10-
10.0.16299
2018-05-04 23:30:06 [scrapy.crawler] INFO: Overridden settings: {'NEWSPIDER_MODULE': 'awaninews.spiders', 'FEED_URI': 'news.csv', 'SPIDER_MODULES': ['awaninews.spiders'
]}
2018-05-04 23:30:06 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.feedexport.FeedExporter',
'scrapy.extensions.logstats.LogStats',
'scrapy.extensions.telnet.TelnetConsole',
'scrapy.extensions.corestats.CoreStats']
2018-05-04 23:30:10 [scrapy.middleware] INFO: Enabled downloader middlewares:
['scrapy.downloadermiddlewares.robotstxt.RobotsTxtMiddleware',
'scrapy.downloadermiddlewares.httputauth.HttpAuthMiddleware',
'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',
'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',
'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',
'scrapy.downloadermiddlewares.retry.RetryMiddleware',
'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',
'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',
'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',
'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',
'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',
'scrapy.downloadermiddlewares.stats.DownloaderStats']
2018-05-04 23:30:10 [scrapy.middleware] INFO: Enabled spider middlewares:
['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',
'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',
'scrapy.spidermiddlewares.referrer.ReferrerMiddleware',
'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',
'scrapy.spidermiddlewares.depth.DepthMiddleware']
2018-05-04 23:30:10 [scrapy.middleware] INFO: Enabled item pipelines:
[]
2018-05-04 23:30:10 [scrapy.core.engine] INFO: Spider opened
2018-05-04 23:30:10 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min), scraped 0 items (at 0 items/min)
2018-05-04 23:30:10 [scrapy.extensions.telnet] DEBUG: Telnet console listening on 127.0.0.1:6023
2018-05-04 23:30:14 [scrapy.core.engine] DEBUG: Crawled (200) <GET http://www.astroawani.com/robots.txt> (referer: None)
2018-05-04 23:30:23 [scrapy.core.engine] DEBUG: Crawled (200) <GET http://www.astroawani.com/search/index/page/6/q/siber> (referer: None)
2018-05-04 23:30:24 [scrapy.core.scraper] DEBUG: Scraped From <200 http://www.astroawani.com/search/index/page/6/q/siber>
{'Title': [u'<div class="row listing-list">\n
<div class="listing-nav col-xs-12">\n
<div class="row listing-nav-row">\n
<div cl
ass="listing-nav-left">\n
<span style="text-transform:none">Carian untuk \'siber\'</span>
</div>\n
<div class="listing-nav-right">\n
<a href="/search?q=siber" class="active">Semua</a><span class="divider"></span><a href="/search/news/q/siber">Berita</a><span class="divider"></span><a href="/search
/photos/q/siber">Foto</a><span class="divider"></span><a href="/search/videos/q/siber">Video</a>
</div>\n
</div>\n
</div>\n
<div class="listing-item col-xs-6 col-sm-4 col-md-4">\n

```

Rajah 3 Proses perlombongan berita menggunakan 'scrapy'

Berita jenayah siber dilombong menggunakan perisian ‘scrapy’ yang menggunakan bahasa pengaturcaraan ‘python’. Jumlah keseluruhan berita yang berjaya dilombong daripada akhbar tempatan online adalah sebanyak 97 berita. Berita dilombong daripada 4 jenis akhbar tempatan online iaitu berita awani, berita harian, utusan dan kosmo. 4 jenis akhbar tempatan online telah digunakan. Hal ini kerana akhbar tempatan online lain mempunyai sedikit masalah. Contohnya laman web harian metro tidak mempunyai satu fail iaitu fail ‘robot.txt’. Ketiadaan fail itu menyebabkan berita tidak dapat dilombong daripada akhbar tempatan tersebut.



The image shows a CSV file with 26 rows of data. Each row contains a single HTML snippet. Rows 1 through 11 contain the tag <div class="listing-text">. Rows 12 through 26 contain the tag <div class="listing-item col-xs-6 col-sm-4 col-md-4">. A large watermark 'Copyright@FTSM' is overlaid diagonally across the entire image.

Rajah 4 Gambaran berita yang berjaya dilombong disimpan di fail berformat ‘csv’

Peringkat seterusnya, berita jenayah siber yang berjaya dilombong akan dikategorikan mengikut beberapa kategori menggunakan kaedah ‘k-means clustering’. Hasil daripada pemerhatian yang dapat saya lakukan, nilai ‘k’ telah ditetapkan kepada 3 dimana berita jenayah siber ini akan dikategorikan kepada 3 kategori. Nilai 3 dipilih kerana terdapat 3 perkataan yang unik dan berita yang dilombong boleh dikategorikan kepada 3 kategori.

Rajah 5 memaparkan 3 kategori yang dapat diperolehi daripada proses ‘k-means clustering’. Selepas proses mengkategorikan berita, statistik jenayah siber akan dibina mengikut bulan. Statistik dibina menggunakan perisian daripada laman web ‘live.amcharts.com’.

```
(base) C:\Users\User\Desktop>python crime.py
Top terms per cluster:
Cluster 0: penggodam kanak data laman dan sosial keselamatan guna web sasar
Cluster 1: siber jenayah serangan di khas ancaman buli dengan jawatankuasa keselamatan
Cluster 2: wang terlibat terdedah individu judi haram mata godam kripto pengganas

Prediction
[2]
[0]
```

Rajah 5 Keputusan proses 'k-means clustering'

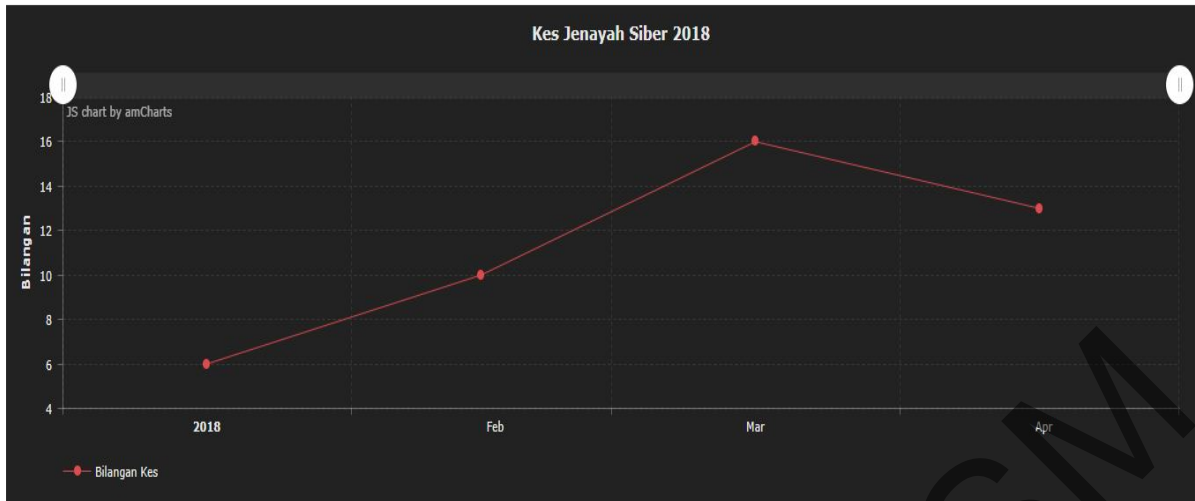
Berita yang berjaya dilombong dan dikategorikan akan dipaparkan ke laman web seperti berikut.

The screenshot shows the CyberCX website interface. At the top, there is a navigation bar with the logo 'CyberCX' on the left and menu items 'BERITA', 'KLASIFIKASI', and 'STATISTIK' on the right. Below the navigation bar is a large blue banner with the text 'CyberCX' and 'Laman web yang memaparkan segala berita berkaitan jenayah siber di Malaysia.' Below the banner, there is a section titled 'BERITA BERKAITAN PENIPUAN/SCAM'. Under this section, there is a list of news items with dates and titles:

- Mac 30, 2018: Editor portal ditangkap sebar berita palsu
- Februari 12, 2018: Hati-hati pilih 'kawan' siber
- Januari 28, 2018: Pengguna di negara ini diingatkan tidak mudah terpengaruh dengan iklan jualan murah yang dipamerkan di laman sosial bagi mengelak menjadi mangsa penipuan siber.

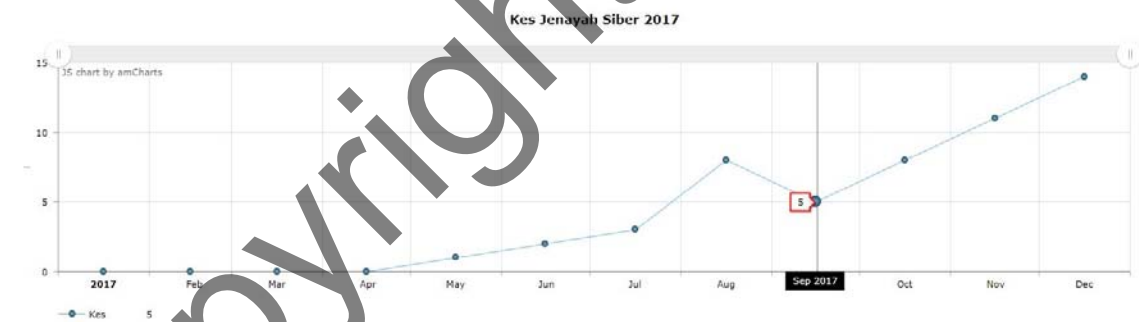
Rajah 6 Senarai berita jenayah siber penipuan

Rajah 7 memaparkan gambaran statistik jenayah siber yang berlaku pada tahun 2018.



Rajah 7 Statistik kes jenayah siber pada tahun 2018

Rajah 8 memaparkan gambaran statistik jenayah siber yang berlaku pada tahun 2017.



Rajah 8 Statistik kes jenayah siber pada tahun 2017

Pengujian sistem perlu dijalankan selepas pembangunan sistem ini untuk memastikan sistem ini dapat berfungsi dengan baik dan menepati spesifikasi yang ditetapkan. Pengujian pertama merupakan proses perlombongan berita daripada akhbar tempatan online. Pengujian ini dianggap berjaya apabila segala berita jenayah siber berjaya dilombong daripada akhbar tempatan online tersebut. Pengujian juga dijalankan ketika proses pengumpulan berita mengikut kategori yang menggunakan kaedah k-means clustering.

6 KESIMPULAN

Laman web jenayah siber ini dijangka dapat memberi kesedaran dan pengetahuan berkaitan kes jenayah siber kepada masyarakat Malaysia. Kemudahan ini memainkan peranan penting di negara yang sedang membangun ini untuk mengurangkan mangsa jenayah siber serta kes jenayah siber di Malaysia.

Penggunaan scrapy dan bootstrap dalam projek ini dapat memudahkan kerja melombong berita dan membangunkan laman web jenayah siber ini. Fungsi yang mesra pengguna dalam perisian scrapy telah memudahkan proses perlombongan berita jenayah siber dari akhbar tempatan online.

7 RUJUKAN

- Boylan, C. 2017. Data Analysis Software and Tools
<http://www.informationbuilders.com/data-analysis> [12 November 2017]
- Brown, M. 2012. Data Mining techniques
<https://www.ibm.com/developerworks/library/ba-data-mining-techniques/>[12 November 2017]
- IBM Knowledge Center. 2017. Clustering Models
https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.modeler.help/nodes_clusteringmodels.htm [9 Oktober 2017]
- Rouse, M. 2017. What is data visualization?
<http://searchbusinessanalytics.techtarget.com/definition/data-visualization> [21 Oktober 2017]
- Syahrul Nizam. 2014. Jenayah Digital daripada penyelidikan kepada penyelesaian. Universiti Malaysia Sarawak (UNIMAS). [11 Oktober 2017]
- WIDESKILLS. 2017. Data Mining Architecture
<http://www.wideskills.com/data-mining-tutorial/data-mining-architecture>[11 Disember 2017]
- WIDESKILLS. 2017. Data Mining Tasks | Data Mining tutorial
<http://www.wideskills.com/data-mining-tutorial/05-data-mining-tasks> [12 November 2017]