

SISTEM PERAMAL DIET MENGGUNAKAN PEMBELAJARAN MESIN RANDOM FOREST

NUR HIDAYAH BINTI MOHD BADARI
TS. DR. SABRINA BINTI TIUN

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,
Selangor Darul Ehsan, Malaysia*

Abstrak

Sistem ini akan meramal kalori bagi pengguna yang ingin menurun atau menambah berat badan mengikut keperluan pengguna. Algoritma yang akan digunakan dalam projek ini adalah salah satu model pembelajaran mesin (*machine learning*) iaitu *Random Forest* dan bahasa yang akan digunakan adalah *Python*. Pengguna akan memasukkan data mereka seperti jantina, umur, berat badan, tinggi dan aktiviti fizikal sebagai input. Sistem ini akan meramal kalori yang diperlukan oleh pengguna dan memberikan cadangan menu mengikut kesesuaian pengguna. Sistem yang sedia ada tidak mempunyai cadangan makanan yang boleh diambil oleh pengguna. Data yang saya gunakan bagi membina sistem ini adalah melalui pembacaan umum daripada laman sesawang *Kaggle. Machine Learning Life-Cycle* merupakan metodologi yang akan digunakan bagi membangunkan Sistem Peramal Diet. Sistem ini akan digunakan oleh individu yang ingin mengamalkan gaya hidup sihat dengan mengawal pengambilan kalori. Antaramuka bagi sistem ini adalah aplikasi web yang dibangunkan dengan menggunakan *Anvil*.

Pengenalan

1.1 Latar Belakang Projek

Kesibukan hidup dan kekangan masa merupakan antara faktor yang menyebabkan seseorang tidak mengawal pemakanan dan jadual menjadi tidak teratur. Kebanyakan golongan muda sering meninggalkan sarapan dan terus mengambil makanan tengah hari kerana tidak mempunyai selera makan atau kekangan masa. Hal ini menyebabkan perancangan diet yang teratur menjadi suatu keperluan dalam kehidupan seharian. Terdapat segelintir golongan yang mengambil jurulatih peribadi bagi mengawal pengambilan makanan mereka supaya dapat kekal sihat. Para penyelidik mengambil idea-idea ini dalam pembinaan sistem yang mana sistem tersebut menggunakan algoritma pembelajaran mesin untuk mengumpul data-data pengguna seperti jantina, umur, berat badan semasa, dan ketinggian. Data-data tersebut akan dianalisis untuk mendapatkan cadangan makanan yang sesuai buat pengguna.

Algoritma pembelajaran mesin banyak digunakan di dalam bidang kesihatan terutamanya dalam pengurusan diet ini. Contoh kajian yang mengaplikasikan pembelajaran mesin dalam bidang ini adalah kajian yang ditulis oleh Ramyaa et al (2019). Kajian ini dibuat hasil daripada penemuan jenis fenotip yang sama di dalam satu populasi wanita dan meramal berat badan mereka. Kumpulan dibahagikan mengikut pemakanan dan latihan fizikal yang dicadangkan kerana perbezaan keputusan daripada individu wanita yang berlainan menyebabkan diet dan senaman yang dicadangkan lebih efektif mengikut kesesuaian individu. Kajian yang dibuat oleh Yin et al (2021) adalah sistem yang merekod data-data di hospital untuk membahagikan pesakit mengikut status nutrisi mereka sama ada ringan, sederhana atau kritikal. Fungsi sistem ini adalah untuk mengesan pesakit yang menghadapi kanser berdasarkan status nutrisi seseorang.

1.2 Objektif

1. Membina satu model peramal diet menggunakan algoritma pembelajaran mesin iaitu *Random Forest*
2. Membina satu laman sesawang yang sesuai bagi mereka yang ingin mendapatkan berat badan yang ideal dengan mencadangkan makanan yang sesuai.
3. Menggunakan set data yang sebenar bagi mendapatkan ramalan yang tepat berdasarkan kepada data yang diperolehi.

1.3 Skop

Skop projek ini merangkumi pelajar institut pengajian tinggi (IPT) yang mengalami masalah dengan diet mereka. Mereka ingin mengamalkan diet yang seimbang dan memasak menu ringkas yang boleh dilakukan di rumah atau kolej sahaja. Faktor yang menyebabkan diet mereka tidak menampakkan perubahan serta semangat yang pada awalnya berkobar-kobar semakin merundum kerana kesibukan kerja dan kekangan masa.

Skop yang lebih tertumpu kepada pelajar IPT ini kerana kebanyakan pelajar IPT sekarang ingin memiliki berat badan yang ideal dan ingin mengamalkan gaya hidup sihat. Mereka tidak mempunyai masa yang cukup untuk menghadiri kelas khusus yang berdekatan untuk mendapatkan jurulatih khas. Mereka juga perlu mengeluarkan kos yang mahal untuk mendaftarkan diri ke kelas khusus dan untuk membayar jurulatih khas.

1.4 Justifikasi dan Kepentingan

Projek Sistem Peramal Diet ini perlu dilakukan bagi memudahkan individu yang ingin mencapai berat badan yang ideal tanpa perlu mengeluarkan wang bagi mendapatkan maklumat seperti nilai kalori yang diperlukan dan cadangan makanan yang sesuai buat pengguna. Kebanyakan sistem dan aplikasi diet yang sedia ada perlu dibayar bagi mendapatkan maklumat asas. Selain daripada itu, Sistem Peramal Diet ini menggunakan model pembelajaran mesin *Random Forest* bagi meramal kalori dimana model ini dapat memberikan ketepatan yang tinggi dan output yang tepat. Hasil daripada projek ini dapat dijadikan sebagai rujukan bagi penyelidik atau pelajar yang menulis tesis berkaitan dengan tajuk ini supaya projek dapat diperbaiki dan ditambahbaik pada masa akan datang.

1.5 Metodologi

Bagi menjayakan projek ini, *Machine Learning Life Cycle* atau kitaran hidup pembelajaran mesin telah dipilih sebagai metodologi. Metodologi ini merangkumi tujuh fasa dimana kesemua fasa ini perlu diikuti mengikut turutan yang telah ditetapkan. Fasa-fasa tersebut adalah perolehan data (*Data Acquisition*), penyediaan data (*Data Preparation*), hipotesis dan pemodelan (*Hypothesis and Modelling*), penilaian dan tafsiran (*Evaluation and Interpretation*), mengimplementasi (*Deployment*), operasi (*Operations*), dan pengoptimuman (*Optimization*). Projek ini memerlukan pengguna untuk memasukkan data mereka seperti jantina, umur, berat badan terkini, tinggi, dan

aktiviti fizikal sebagai input untuk meramal kalori yang diperlukan oleh pengguna. Selain itu, sistem juga akan menunjukkan cadangan makanan seimbang yang boleh diambil tidak kira sama ada pengguna ingin menurunkan berat badan, menambah berat badan atau ingin mengamalkan gaya hidup yang sihat. Antaramuka bagi sistem ini adalah laman sesawang dan sistem ini bekerja menggunakan mesin pembelajaran yang akan membuat ramalan dan analisis menggunakan data-data pengguna seperti yang dinyatakan di atas.

1.6 Organisasi Laporan

Struktur organisasi laporan teknik ini dimulakan dengan pengenalan bagi memberi gambaran umum tentang projek, tujuan dan proses pelaksanaan pembangunan Sistem Peramal Diet. Justifikasi dan kepentingan pembangunan sistem ini juga merupakan salah satu isi yang terdapat pada bahagian pengenalan bagi menerangkan tujuan dan manfaat yang diperolehi daripada kajian ini. Sorotan susastera bertujuan membuat perbandingan antara laman sesawang yang sedia ada dan sistem yang ingin dibangunkan. Selain itu, perbandingan beberapa model pembelajaran mesin juga dibuat bagi memilih model dengan ketepatan yang tinggi dan yang paling sesuai digunakan bagi Sistem Peramal Diet ini. Metodologi kajian adalah penjelasan mengenai proses dan langkah-langkah yang dibuat bagi membangunkan algoritma model peramal diet. Hasil kajian seperti ketepatan dan output yang dikeluarkan akan diterangkan pada bahagian keputusan dan perbincangan. Keseluruhan kajian disimpulkan dan pencapaian objektif dinyatakan di dalam kesimpulan dimana implikasi, kelemahan dan cadangan penambahbaikan juga disertakan sekali pada bahagian terakhir ini.

1.7 Sorotan Susastera

Perbandingan sistem yang digunakan adalah '*Ideal Weight Calculator*' yang dibina pada tahun 2008 dan '*Weight Loss Predictor Calculator*' yang dibina oleh Robert S. Lasater dan James K. Starling. Kedua-dua sistem ini dibangunkan di platform laman sesawang yang boleh diakses di mana-mana gajet seperti telefon pintar, tablet atau komputer riba. Sistem '*Ideal Weight Calculator*' memberikan pengeluaran yang agak ringkas dan kurang menarik kerana sistem ini hanya memberitahu formula yang digunakan untuk mengira berat badan yang ideal pengguna manakala pengeluaran bagi sistem '*Weight Loss Predictor Calculator*' terlalu ringkas dan tiada maklumat yang mendalam berkenaan dengan graf yang ditunjukkan. Kedua-dua sistem ini tiada gambar rajah atau peta minda sebagai ringkasan kepada analisis yang dibuat oleh sistem-sistem ini.

Sistem '*Ideal Weight Calculator*' memberikan penerangan secara am perkaitan antara umur, jantina, ketinggian dengan pengeluaran yang diberikan iaitu formula dan berat badan yang ideal. Sistem ini juga memberikan penerangan berkenaan formula-formula yang dikeluarkan dimana semua formula itu mempunyai format yang sama iaitu berat asas yang diberi ketinggian 5 kaki dengan satu set peningkatan berat badan yang ditambah satu inci untuk setiap ketinggian yang melebihi 5 kaki. Sistem '*Weight Loss Predictor Calculator*' pula menggunakan model keseimbangan tenaga matematik dinamik (*dynamic mathematical energy balance*) yang meramal perubahan berat badan. Sistem ini telah dibangunkan berasal daripada persamaan keseimbangan tenaga yang dirujuk dalam hukum pertama termodinamik. Sistem ini menerangkan model yang digunakan bagi membuat analisis dengan menerangkan formula-formula yang berkaitan dengan model tersebut.

Sistem	<i>Ideal Weight Calculator</i>	<i>Weight Loss Predictor Calculator</i>	Sistem Peramal Diet
<i>Input</i>	Umur, jantina, ketinggian	Jantina, unit, berat semasa, umur, ketinggian, tempoh masa, pengurangan kalori	Jantina, umur, berat badan terkini, berat badan ideal, tinggi dan aktiviti fizikal
<i>Output</i>	Formula, ideal berat badan	Hari, berat badan yang perlu dicapai	Cadangan makanan untuk sarapan, makan tengah hari dan makan malam
Cadangan menu makanan	Tiada cadangan makanan	Tiada cadangan makanan	Mempunyai cadangan makanan mengikut kesesuaian pengguna
Antaramuka	Terlalu kompleks	Terlalu ringkas	Ringkas tetapi padat

Jadual 1.7.1 : Perbandingan sistem yang digunakan di dalam kajian

Algoritma pembelajaran mesin yang digunakan bagi membuat perbandingan adalah *Linear Regression*, *Least Absolute Shrinkage and Selection Operator (LASSO)* dan *Random Forest*. *Linear Regression* adalah model yang menggunakan pendekatan linear yang mengaitkan hubungan antara respons skalar dengan satu atau lebih pembolehubah penjelasan. LASSO merupakan model yang ditambahbaik daripada *Linear Regression* dengan menggunakan penalti L1 dalam memilih atribut yang bersesuaian dan menghilangkan atribut yang tidak relevan untuk membuat latihan model. *Random Forest* pula adalah gabungan daripada beberapa pokok keputusan dan membuat ramalan daripada undian majoriti. Perbandingan algoritma ini dilakukan oleh Gudiwada dan Torati dengan tajuk *Comparing Machine Learning Algorithms for Predicting Calories Burned* yang diterbitkan pada Mac 2023. Daripada ketiga-tiga algoritma ini, model *Random Forest*

menunjukkan ketepatan yang paling tinggi dan paling sesuai digunakan bagi Sistem Peramal Diet ini.

Algoritma	<i>Linear Regression</i>	<i>LASSO</i>	<i>Random Forest</i>
Penerangan	Model yang digunakan untuk membuat ramalan berdasarkan hubungan linear antara input dan output	Penambahbaikan daripada Linear Regression dan menggunakan penalti L1 dalam memilih maklumat yang diperlukan dan menghilangkan maklumat yang kurang/tidak relevan	Model yang menggabungkan beberapa pokok keputusan untuk membuat ramalan dan memilih keputusan yang diundi secara majoriti
Kelebihan	Mudah difahami, tempoh masa yang diambil bagi melatih model adalah cepat dan singkat	Boleh mengatasi overfitting, pemilihan fitur secara automatik	Boleh mengatasi overfitting, boleh digunakan untuk pengkelasan atau regresi
Kekurangan	Boleh digunakan bagi hubungan yang linear sahaja, tidak dapat mengatasi overfitting	Cenderung menghasilkan model yang bias, susah untuk membuat anggaran ralat	Memerlukan lebih banyak data untuk membina pokok yang banyak, tempoh masa yang lama bagi melatih model
Ketepatan	96% - 97%	96%	97% - 99%

Jadual 1.7.2 : Perbandingan algoritma pembelajaran mesin yang digunakan di dalam kajian

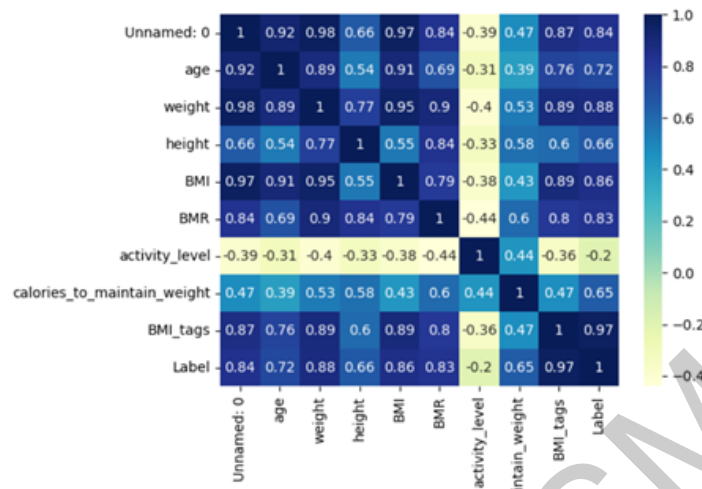
Metodologi Kajian

2.1 Perolehan Data

Perolehan dan pengumpulan data dilakukan semasa proses ini. Data yang digunakan bagi projek ini diperolehi daripada sumber terbuka iaitu *Kaggle* yang mempunyai lebih daripada 10,000 data terkumpul. Data yang diperolehi ini berkaitan dengan pengiraan kalori dimana atribut yang terdapat di dalam set data ini adalah umur, berat, tinggi, jantina, BMI, BMR, aktiviti dan kalori yang perlu dikekalkan. Set data ini telah digunakan oleh 2 orang pengaturcara bagi membina sistem cadangan diet iaitu Kishore Harshan Kumar dan Muhammad Usman. Set data ini sesuai digunakan bagi membangunkan Sistem Peramal Diet kerana atribut yang terdapat di dalam set data ini sepadan dengan *input* dan *output* sistem.

2.2 Penyediaan Data

Data yang diperolehi akan dibersihkan bagi memudahkan proses penilaian dan implementasi. Pra-pemprosesan dilakukan berdasarkan kepada *Exploratory Data Analysis* (EDA) iaitu analisis yang digunakan untuk melihat data yang melebihi keperluan model dan membantu menentukan teknik statistik yang bersesuaian. Pra-pemprosesan yang pertama dilakukan bagi pembinaan projek Sistem Peramal Diet ini adalah membuang data yang tidak diperlukan seperti '*Unnamed: 0*', '*Label*' dan '*BMI_tags*' dengan memanggil pakej *drop()*. Proses ini dilakukan dengan merujuk kepada *heatmap* set data yang digunakan seperti yang ditunjukkan pada gambar rajah di bawah. *Heatmap* tersebut menunjukkan 3 data yang dibuang tersebut merupakan penyumbang terbesar kepada ramalan kalori.



Rajah 2.2.1 : *Heatmap* bagi set data ramalan kalori

Seterusnya, jantina ditukar kepada 0 dan 1 dengan menggunakan *replace()* dimana F yang mewakili 'Female' merupakan 0 manakala M yang mewakili 'Male' merupakan 1. *Data Binning* dan *Label Encoding* juga dilakukan semasa proses pembersihan data. *Data Binning* adalah teknik pra-pemprosesan data yang digunakan untuk mengecilkan ralat daripada kesan pemerhatian. Nilai data yang asal dibahagikan kepada bahagian-bahagian yang kecil dipanggil sebagai *bins* dan nilai tersebut diganti dengan nilai yang telah dikira bersesuaian dengan *bins* tersebut. *Label Encoding* pula adalah teknik yang digunakan bagi menukar data kepada data berangka supaya dapat digunakan oleh model pembelajaran mesin yang hanya menerima data berangka bagi membuat ramalan.

2.3 Hipotesis dan Pemodelan

Fasa hipotesis dan pemodelan dilakukan dengan melalui proses penulisan kod untuk membuat analisis. Set data yang dipilih akan dilatih dengan beberapa algoritma pembelajaran mesin dan memilih algoritma model yang memberikan keputusan yang terbaik iaitu *Random Forest*. Bahasa yang digunakan bagi menulis kod Sistem Peramal Diet ini adalah Python.

2.4 Penilaian dan Tafsiran

Set data akan dibahagikan kepada dua kategori di dalam proses ini dimana kategori tersebut adalah data yang akan dilatih dan data yang akan diuji. Nisbah bagi pembahagian data ini adalah 80% data yang dilatih dan 20% selebihnya digunakan sebagai data yang akan diuji.

2.5 Implementasi

Model *Random Forest* yang digunakan bagi membuat ramalan kalori diimplementasikan kepada aplikasi web dengan menggunakan tools yang dipanggil *Anvil*. Aplikasi web ini akan mengumpul maklumat pengguna yang diperlukan seperti umur, jantina, berat badan, ketinggian dan aktiviti fizikal. Kemudian, maklumat tersebut akan dihantar ke *Google Colab* bagi membuat ramalan dan mendapatkan cadangan menu yang bersesuaian dengan kalori pengguna dimana data-data ini akan diklasifikasikan. Klasifikasi data tersebut akan dihantar semula ke aplikasi web untuk dipaparkan kepada pengguna.

Diet Predictor

Age : 23

Gender : Male Female

Current Weight (kg) : 55

Height (m) : 1.75

Physical Activities : Sedentary Light Moderate Active Extra

Physical Activities Explanation :

Sedentary = little or no exercise
Light = light exercise/sports 1-3 days/week
Moderate = moderate exercise/sports 3-5 days/week
Active = hard exercise/sports 6-7 days a week
Extra = very hard exercise/sports & physical job or 2x training

PREDICT

Rajah 2.5.1 : Antaramuka Sistem Peramal Diet selepas *input* dimasukkan

The calories that should be maintained is 1439.62

Menu Suggestion :

(Monday):	Food	Quantity (g)	
361	Yogurt, nonfat, fruit variety	150.0	
401	Pork, raw, with added solution, separable lean...	124.0	
794	Borage, raw	150.0	
868	Snacks, cheese-flavor, potato chips	67.0	
878	Plums, solids and liquids, extra heavy syrup p...	150.0	
1126	KEEBLER, Club & Cheddar Sandwich Crackers	36.0	
1173	Beef, braised, cooked, choice, trimmed to 1/8"...	150.0, Tuesday:	Food Quantity (g)
363	Corn, raw, yellow, sweet	150.0	
550	Bread, french or vienna (includes sourdough)	101.0	
789	Seeds, paste, sesame butter	25.0	
796	Turnips, unprepared, frozen	43.0	
1236	Broccoli, with salt, drained, boiled, cooked	150.0, Wednesday:	Food Quantity (g)
452	Soy milk, vitamins A and D, with added calcium...	150.0	
833	Alcoholic beverage, canned, tequila sunrise	18.0	
1010	Bread, refrigerated, commercially prepared, wh...	150.0	
1128	Tofu, prepared with calcium sulfate, dried-fro...	31.0	
1135	SUNSHINE, Mozzarella Crackers, CHEEZ-IT	6.0, Thursday:	Food Quantity (g)
237	Oranges, Florida, raw	150.0	
254	Yogurt, fortified with vitamin D, nonfat milk...	150.0	
261	Corn, solids and liquids, regular pack, whole ...	94.0	
856	Sourdock, young leaves (Alaska Native)	37.0	
1083	CAMPBELLS Soup on the Go, Vegetable with Mini...	46.0	
1147	Arrowhead, raw	150.0, Friday:	Food Quantity (g)
467	Beef, raw, select, trimmed to 1/8" fat, separa...	150.0	
495	Cereals, without salt, microwaved, cooked with...	104.0	
599	Milk, sweetened, condensed, canned	150.0	
829	Nuts, unpeeled, dried, european, chestnuts	57.0	
1241	KEEBLER, KEEBLER Chocolate Graham SELECTS	97.0, Saturday:	Food Quantity (g)
15	Lamb, roasted, cooked, choice, trimmed to 1/4"...	150.0	
337	Beef, grilled, cooked, select, trimmed to 1/8"...	51.0	
448	SILK Unsweetened, soy milk	150.0	
479	Infant Formula, ready-to-feed, Infant, Premium...	150.0	
782	Beef, raw, 75% lean meat / 25% fat, ground	150.0	
828	Beans, solids and liquids, seasoned, all style...	23.0	
979	Bagels, toasted, with calcium propionate (incl...	150.0	
1108	Cabbage, without salt, drained, boiled, cooked...	6.0, Sunday:	Food Quantity (g)
53	Snacks, peanut butter, hard, granola bars	98.0	
78	Sesbania flower, with salt, steamed, cooked	150.0	
518	Veal, cooked, seam fat only	150.0	
551	Veal, roasted, cooked, separable lean only, loin	31.0	
738	Pork, unheated, separable lean and fat, bone-l...	150.0	
739	Tofu, prepared with calcium sulfate, fried	118.0	
1030	Game meat, broiled, cooked, 1" steak, separabl...	150.0	
1047	Margarine-like, with salt, approximately 37% f...	150.0	
1143	Fish, smoked and canned (Alaska Native), chino...	150.0	

Rajah 2.5.2 : Antaramuka Sistem Peramal Diet selepas *output* dikeluarkan

Keputusan dan Perbincangan

3.1 Keputusan

Masa yang diambil bagi menyelesaikan ujian adalah 5 ke 6 saat dengan ketepatan 99.97%. Selain daripada itu, senarai menu yang dikeluarkan juga merupakan kriteria bagi menentukan fasa pengujian telah tamat.

0.9997087030009596
 Accuracy score without standardization: 43.13733311650251
 0.9997073905529194
 Accuracy score with standardization: 43.33168976452772

Rajah 3.1.1 : Ketepatan bagi model *Random Forest* meramal kalori

1125	KELLOGG'S, Frosted Chocolate Baked Bites, POP-...	150.0,	'Wednesday':	Food	Quantity (g)
63	Snacks, all flavors, QUAKER OATMEAL TO GO, gra...	150.0			
249	Fruit salad, solids and liquids, juice pack, c...	150.0			
367	CAMPBELL'S Soup on the Go, Vegetable Beef Soup	29.0			
374	Pork, roasted, heated, separable lean and fat,...	150.0			
1012	KASHI, Frozen Entree, Roasted Garlic Chicken F...	150.0			
1123	Beverages, Diet Fruit Medley, V8 SPLASH Juice ...	150.0			
1125	Turkey, roasted, with added solution, meat onl...	18.0			
1220	Cheese, cheshire	1.0,	'Thursday':	Food	Quantity (g)
71	ARCHWAY Home Style Cookies, Dutch Cocoa	150.0			
318	Pork, pan-fried, cooked, separable lean only, ...	150.0			
339	Beverages, whey protein powder isolate	150.0			
640	wheat, cooked, KAMUT khorasan	16.0			
810	SILK Plain, soy milk	150.0			
817	Ice cream bar, covered with chocolate and nuts	94.0			
1060	Beverages, powder, regular, instant, coffee	150.0			
1130	Babyfood, pretzels	31.0			
1136	Pretzels, unsalted, soft	150.0,	'Friday':	Food	Quantity (g)
478	Gravy, ready-to-serve, canned, turkey	150.0			
511	Sausage, links, sweet, Italian	150.0			
624	Beef, raw, select, trimmed to 1/8" fat, separa...	150.0			
757	Infant formula, prepared from powder, NEXT STEP...	150.0			
779	Snacks, with peanuts, caramel-coated, popcorn	117.0			
1156	Cowpeas, raw, leafy tips	150.0			
1221	Beef, braised, cooked, all grades, trimmed to ...	115.0			
1239	Beef, raw, choice, trimmed to 0" fat, separabl...	150.0,	'Saturday':	Food	Quantity (g)
257	MORNINGSTAR FARMS California Turk'y Burger, un...	150.0			

Rajah 3.1.2 : Senarai menu setelah fasa pengujian selesai

3.2 Analisis Keputusan

Terdapat sedikit perbezaan ketepatan kerana projek ini menggunakan *Random Forest Regressor*. Algoritma yang berkaitan dengan pokok keputusan (*Decision Tree*) tidak perlu dilakukan normalisasi data (*Data Normalization*) tetapi bagi *regressor*, algoritma akan lebih dipengaruhi oleh nilai yang lebih *high-end* jika data tidak diubah. Ini bermakna bahawa algoritma ini mungkin akan lebih tepat dalam meramalkan nilai tinggi daripada nilai rendah. Akibatnya, transformasi seperti log-transform akan mengurangkan kepentingan relatif nilai tinggi ini, justeru membuat generalisasi dengan lebih baik. Daripada keputusan diatas, algoritma Random Forest tanpa normalisasi dipilih bagi projek ini kerana ia mempunyai ketepatan yang lebih tinggi.

3.3 Perbandingan Kajian

Kajian yang dibuat oleh Gudiwada Vijayalakshmi dan Torati Sridurga menunjukkan bahawa model Random Forest dapat meramal kalori dengan ketepatan 97% manakala ketepatan yang diperolehi oleh Devhadvani adalah 99.78%. Hasil kajian yang diperolehi daripada projek Sistem Peramal Diet ini adalah 99.97%. Hasil kajian projek ini mempunyai persamaan dengan kajian yang dibuat oleh Devhadvani kerana kedua-dua kajian menggunakan model Random Forest Regressor dan nisbah pembahagian data latihan dan data ujian adalah sama iaitu 80:20.

3.4 Penjelasan

Sistem Peramal Diet menggunakan model *Random Forest* untuk membuat ramalan dan set data yang digunakan adalah set data yang relevan dan sesuai digunakan di dalam kajian ini. Set data yang diperolehi daripada *Kaggle* ini terdiri daripada atribut-atribut yang diperlukan bagi meramal kalori seperti berat badan, umur, jantina, tinggi dan aktiviti fizikal. Pra-pemprosesan dan pembersihan data dilakukan dengan menggunakan *Exploratory Data Analysis* (EDA) bagi meningkatkan kualiti model dan sistem. Model *Random Forest* meramal kalori yang diperlukan oleh pengguna berdasarkan kepada atribut set data yang diambil dan input yang diperolehi daripada pengguna. Hasil ramalan ini dapat membantu pengguna menambah atau mengurangkan berat badan dengan senarai cadangan menu yang diberikan.

3.5 Implikasi

Sistem Peramal Diet ini dapat membantu individu mengamalkan gaya hidup yang sihat dan mengawal pemakanan mereka. Pengguna dapat memahami keperluan kalori harian mereka dengan lebih baik dan membantu mengurangkan risiko penyakit seperti obesiti. Daripada ramalan kalori tersebut, senarai cadangan makanan yang dikeluarkan dapat membantu pengguna mengatur dan mengawal pengambilan makanan seharian mereka. Sistem ini juga dapat meningkatkan kesedaran masyarakat tentang kepentingan pengambilan makanan yang seimbang dan amalan gaya hidup sihat. Kajian ini dapat dijadikan sebagai rujukan bagi penyelidik atau pelajar yang menulis tesis berkaitan dengan tajuk ini supaya projek dapat diperbaiki dan ditambahbaik pada masa akan datang.

3.6 Kelemahan dan Cadangan Penambahbaikan

Kelemahan bagi Sistem Peramal Diet ini adalah set data yang *overfitting* dengan ketepatan 100%. Setelah mencari beberapa cadangan penyelesaian bagi masalah ini, cara yang paling sesuai bagi pra-pemprosesan data adalah *Exploratory Data Analysis* dan *Data Binning*. Kedua-dua cara ini

membantu model *Random Forest* mengatasi masalah *overfitting* tersebut dan ramalan dapat dibuat dengan baik dan tepat. Sistem Peramal Diet ini dibangun dengan menggunakan *Google Colab* sebagai platform bagi penulisan kod dan *Anvil* sebagai platform bagi antaramuka. Kod yang ditulis di dalam *Google Colab* akan diputuskan sambungan selepas jangka masa yang tertentu. Hal ini menyebabkan pengguna tidak dapat menggunakan Sistem Diet Peramal pada setiap masa dalam tempoh 24 jam.

Cadangan penambahbaikan bagi projek ini adalah menggunakan teknik *Cross-Validation* di dalam pra-pemprosesan. Selain daripada *Exploratory Data Analysis* dan *Data Binning*, teknik ini juga dapat membantu model mengatasi *overfitting* dan dapat membuat ramalan dengan lebih baik ketika mendapat *input* yang baru atau maklumat yang tidak terdapat di dalam set data semasa membuat latihan. Cadangan bagi menambahbaik *backend* dan *frontend* Sistem Peramal Diet adalah menggunakan perisian atau *software* yang tidak terputus sambungan seperti *Pycharm* atau *Eclipse* sebagai penulisan kod algoritma model. Bagi antaramuka sistem pula, perisian seperti *Streamlit* atau *Flask* boleh digunakan untuk mengatasi kekangan yang dihadapi.

Kesimpulan

Sistem Peramal Diet ini menggunakan algoritma *Random Forest* untuk meramal keperluan kalori pengguna berdasarkan set data yang diperolehi dari *Kaggle*. Maklumat yang diterima daripada pengguna sebagai input adalah umur, jantina, berat badan, tinggi dan aktiviti fizikal. Sistem ini akan memberikan ramalan jumlah kalori yang diperlukan oleh pengguna serta menyediakan senarai cadangan menu yang sesuai mengikut keperluan kalori tersebut. Metodologi yang digunakan di dalam projek ini adalah *Machine Learning Life Cycle* dimana fasa yang terlibat adalah perolehan data, penyediaan data, hipotesis dan pemodelan, penilaian dan tafsiran, serta implementasi. Hasil akhir pengujian menunjukkan algoritma *Random Forest* dapat memberi

ketepatan yang tinggi iaitu 99.97% dan dapat mengeluarkan *output* yang dijangka dengan baik. Sistem Peramal Diet dapat menerima *input* yang diperlukan daripada pengguna dan dapat mengeluarkan *output* yang bersesuaian iaitu nilai kalori yang diramal dan senarai cadangan makanan yang diperlukan oleh pengguna. Objektif utama bagi kajian ini adalah membina satu model peramal diet menggunakan algoritma pembelajaran mesin iaitu *Random Forest* dan objektif ini telah dicapai. Sistem Peramal Diet ini dapat membantu individu mengamalkan gaya hidup yang sihat dan mengawal pengambilan makanan mereka. Implikasi dari sistem ini juga dapat meningkatkan kesedaran masyarakat tentang kepentingan menjaga kesihatan dan memahami keperluan kalori seharian mereka. Sistem ini dibangunkan menggunakan *Google Colab* dimana platform ini mempunyai tempoh masa yang terhad dan menyebabkan pengguna tidak dapat mengakses sistem pada setiap masa. Cadangan penambahbaikan bagi masalah ini adalah menggunakan perisian seperti *PyCharm* atau *Eclipse* bagi memudahkan pengguna mengakses sistem. Dengan mengimplimentasikan cadangan ini, Sistem Peramal Diet menjadi lebih efisien dan dapat digunakan dengan lebih baik oleh pengguna.

Penghargaan

Bersyukur ke hadrat Ilahi dengan limpahan kurnia-Nya yang dianugerahkan kepada saya untuk menyiapkan laporan teknikal ini dengan jayanya. Terlebih dahulu saya ingin mengucapkan ribuan terima kasih kepada penyelia saya iaitu Dr. Sabrina Tiun di atas bimbingan, sokongan, nasihat, teguran dan kesabaran yang diberikan sepanjang menyelia saya dalam melaksanakan projek tahun akhir ini.

Selain itu, saya juga ingin mengucapkan terima kasih yang tidak terhingga kepada keluarga saya kerana memberi sokongan moral yang berterusan sepanjang penulisan laporan teknikal ini. Mereka banyak memberi kata-kata semangat supaya saya dapat terus bangkit bagi meneruskan penulisan saya.

Seterusnya, saya turut ingin mengucapkan jutaan terima kasih kepada rakan-rakan saya kerana banyak memberi pendapat dan cadangan penambahbaikan untuk laporan ini. Mereka juga banyak memberi peringatan disaat saya alpa sepanjang proses penulisan laporan teknikal ini. Mereka membantu saya dengan menjawab setiap pertanyaan dan persoalan yang diutarakan kepada mereka.

Akhir kata, saya ingin mengucapkan terima kasih kepada semua yang terlibat secara langsung atau tidak langsung dalam penulisan laporan projek tahun akhir saya ini. Terima kasih.

RUJUKAN

- Avik, D. 2023. *Random Forest Regression in Python*.
<https://www.geeksforgeeks.org/random-forest-regression-in-python/> [5 Jun 2023].
- Calculator.net. (2018). Ideal Weight Calculator.
<https://www.calculator.net/ideal-weight-calculator.html?ctype=metric&age=21&sex=f&heightfeet=5&heightinch=10&heightmeter=155&printit=0&x=117&y=17>
- Devhadvani. (2021). *Calorie Prediction Machine Learning Project using Python*.
<https://devhadvani147.medium.com/calorie-prediction-machine-learning-project-using-python-b85a9dfc9220> [24 May 2021].
- Gudiwada, V. Torati, S. (2023). *Comparing Machine Learning Algorithms for Predicting Calories Burned*. <https://www.jetir.org/papers/JETIR2303465.pdf> [Mac 2023].
- Kirk et al. (2022, 27 September). *Machine Learning in Nutrition Research*. Oxford Academic. [https://academic.oup.com/advances/advance-article/doi/10.1093/advances/nmac103/6724380?searchresult=1&utm_source=TrendMD&utm_medium=cpc&utm_content=Advances in Nutrition 1&utm_campaign=Advances in Nutrition TrendMD 1&login=false](https://academic.oup.com/advances/advance-article/doi/10.1093/advances/nmac103/6724380?searchresult=1&utm_source=TrendMD&utm_medium=cpc&utm_content=Advances%20in%20Nutrition%201&utm_campaign=Advances%20in%20Nutrition%20TrendMD%201&login=false)
- Ramyaa et al. (2019). *Phenotyping Women Based on Dietary Macronutrients, Physical Activity, and Body Weight Using Machine Learning Tools*. <https://www.mdpi.com/2072-6643/11/7/1681>
- Robert S. Lasater dan James K. Starling. Weight Loss Predictor Calculator. Pennington Biomedical Research Center. <https://www.pbrc.edu/research-and-faculty/calculators/weight-loss-predictor/>
- Sidhantha, P. 2023. *Binning in Data Mining*.
<https://www.geeksforgeeks.org/binning-in-data-mining/> [11 Januari 2023].
- Sruthi, E. 2021. *Understand Random Forest Algorithms With Examples*.
<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/> [17 Jun 2021].
- Yin et al. (2021). *A fusion decision system to identify and grade malnutrition in cancer patients: Machine learning reveals feasible workflow from representative real-world data*. Science Direct.
<https://www.sciencedirect.com/science/article/abs/pii/S0261561421003241>

Nur Hidayah Binti Mohd Badari (A181124)
 Ts. Dr. Sabrina Binti Tiun
 Fakulti Teknologi & Sains Maklumat,
 Universiti Kebangsaan Malaysia