

SISTEM PENCADANGAN MAKANAN DIPERIBADIKAN MENGUNAKAN TEKNIK PENYARINGAN BERDASARKAN KANDUNGAN

NURIN JAZLINA BINTI FEROZ

DR. HAZURA MOHAMED

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

ABSTRAK

Projek ini membangunkan sebuah sistem cadangan makanan berasaskan kesihatan menggunakan pendekatan *Content-Based Filtering* (CBF) yang disokong oleh model pembelajaran mesin bagi menyediakan cadangan makanan mengikut keperluan kalori individu. Keperluan kalori dikira menggunakan formula *Harris-Benedict* berdasarkan input pengguna seperti umur, berat badan, tinggi, jantina, dan tahap aktiviti fizikal. Dataset yang digunakan terdiri daripada 1174 makanan dan 59 ciri nutrien yang telah melalui proses pembersihan termasuk penyingkiran nilai hilang, penyeragaman teks dan penapisan makanan tidak halal. Model ramalan kalori dilatih menggunakan algoritma regresi seperti *Linear Regression*, *Decision Tree*, dan *Random Forest*, dan dinilai berdasarkan metrik seperti RMSE (*Root Mean Squared Error*), R^2 (*R-squared*) dan NRMSE (*Normalized Root Mean Squared Error*). Hasil perbandingan menunjukkan bahawa model *Random Forest Regressor* memberikan prestasi terbaik dan digunakan sebagai model utama sistem. Sistem ini dibangunkan menggunakan platform Gradio yang membolehkan pengguna berinteraksi secara langsung dengan antara muka, memasukkan data dan menerima kombinasi tujuh makanan yang dicadangkan berdasarkan keperluan kalori harian mereka. Sistem ini bukan sahaja memudahkan pengguna membuat pilihan makanan yang lebih sihat, tetapi juga berpotensi untuk dikembangkan dalam aplikasi sebenar sebagai alat sokongan gaya hidup sihat yang berasaskan data dan berfokus kepada individu.

PENGENALAN

Projek ini bertujuan membangunkan Sistem Pencadangan Makanan Diperibadikan yang menggunakan teknik Penyaringan Berdasarkan Kandungan (Content-Based Filtering, CBF). Dalam era digital dan gaya hidup moden, ramai individu sukar membuat pilihan makanan yang sesuai dengan keperluan dan gaya hidup masing-masing. Sistem ini dibangunkan untuk membantu pengguna mengenal pasti makanan yang selaras dengan profil peribadi mereka, dengan harapan dapat menyokong gaya pemakanan yang lebih bijak.

Sistem ini berfungsi dengan menerima input pengguna seperti umur, jantina, berat, tinggi, dan tahap aktiviti fizikal. Data ini digunakan untuk mengira keperluan kalori harian melalui formula BMR dan TDEE. Berdasarkan hasil pengiraan tersebut, sistem akan mencadangkan makanan yang sesuai dengan keperluan individu menggunakan pendekatan CBF, yang membandingkan ciri pengguna dengan ciri makanan dalam dataset. Ini menjadikan setiap cadangan bersifat diperibadikan.

Menurut laporan WHO (*World Health Organization*, 2019), pemakanan tidak sihat adalah punca utama penyakit kronik seperti obesiti dan diabetes, yang menyumbang kepada berjuta-juta kematian setiap tahun. Dengan menggabungkan pembelajaran mesin dan pemprosesan data nutrisi, sistem ini memperlihatkan cara teknologi boleh digunakan secara praktikal dalam kehidupan seharian. Dalam masa yang sama, ia turut menunjukkan potensi aplikasi sains data dalam menyelesaikan isu pemilihan makanan yang sesuai. Cadangan makanan yang dihasilkan dijangka lebih tepat dan relevan dengan keadaan pengguna berbanding kaedah tradisional.

Projek ini dibangunkan menggunakan bahasa pengaturcaraan Python dan antaramuka pengguna dibina dengan Gradio bagi memastikan interaksi yang mudah dan mesra pengguna. Sistem ini direka tanpa ciri log masuk, menjadikannya lebih ringkas. Secara keseluruhan, sistem ini menunjukkan cara pendekatan diperibadikan dalam sistem cadangan boleh menyumbang kepada gaya hidup yang lebih teratur.

METODOLOGI KAJIAN

Metodologi kajian ini merangkumi beberapa fasa utama bermula daripada pemerolehan data, pembersihan data, pembangunan model, sehingga penjana cadangan makanan kepada pengguna. Proses pembangunan sistem ini dilaksanakan secara berperingkat, dengan penekanan kepada ketepatan pemprosesan data dan keberkesanan model yang dibangunkan.

Pengumpulan Data

Data yang digunakan dalam kajian ini diperoleh daripada set data makanan yang mengandungi maklumat pemakanan, kuantiti, dan kategori makanan. Set data ini mengandungi 1,175 entri dengan 61 ciri nutrien yang berbeza yang mengandungi maklumat nutrisi seperti kalori, protein, lemak, karbohidrat, vitamin, mineral, dan sebagainya.

Pembersihan Data

Proses pra-pemprosesan dilakukan untuk memastikan kualiti data berada pada tahap optimum sebelum analisis dan pembinaan model dijalankan. Langkah-langkah yang dilaksanakan merangkumi pembuangan data pendua menggunakan `pandas.drop_duplicates()`, pengisian nilai kosong (missing values) dengan 0.00 menggunakan `pandas.fillna()`, penukaran format teks ke huruf kecil bagi lajur kategori seperti nama makanan, jenama, dan kategori menggunakan `.str.lower()`, penyahsenaraian data tidak halal bagi memastikan sistem sesuai digunakan oleh pengguna Muslim, serta penukaran jenis data (data type casting) untuk memastikan nilai berangka berada dalam format float atau int yang betul. Keseluruhan proses pra-pemprosesan ini dilaksanakan menggunakan pustaka Python seperti Pandas dan NumPy.

Pembangunan Model

Model pembelajaran mesin berasaskan *Content-Based Filtering* dengan pendekatan regresi dibangunkan untuk mengira keperluan kalori harian pengguna. Input model merangkumi data berat, tinggi, umur, jantina, dan tahap aktiviti fizikal pengguna. Hasil pengiraan kalori ini kemudiannya digunakan sebagai asas untuk mencadangkan senarai makanan yang sesuai.

Reka Bentuk dan Pembangunan Antara Muka Sistem

Antara muka dibangunkan menggunakan Python dan Gradio, membolehkan pengguna memasukkan data yang diperlukan untuk pengiraan kalori dan seterusnya memaparkan senarai cadangan makanan. Reka bentuk fokus kepada kesederhanaan penggunaan dan paparan hasil yang jelas.

Pengujian Model

Model diuji menggunakan subset data untuk menilai ketepatan ramalan kalori dan kesesuaian cadangan makanan yang dihasilkan. Pemilihan cadangan dinilai berdasarkan kesepadanan nilai kalori dan ciri nutrien dengan keperluan pengguna. Penilaian prestasi model regresi menggunakan Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE), dan R-squared (R^2).

Penjanaan Cadangan Makanan

Berdasarkan output model, sistem menjana senarai makanan yang memenuhi julat kalori dan profil nutrien yang sesuai untuk pengguna. Senarai ini dipaparkan melalui antara muka sistem yang dibangunkan, membolehkan pengguna melihat cadangan secara terus setelah mengisi maklumat diri mereka.

KEPUTUSAN DAN PERBINCANGAN

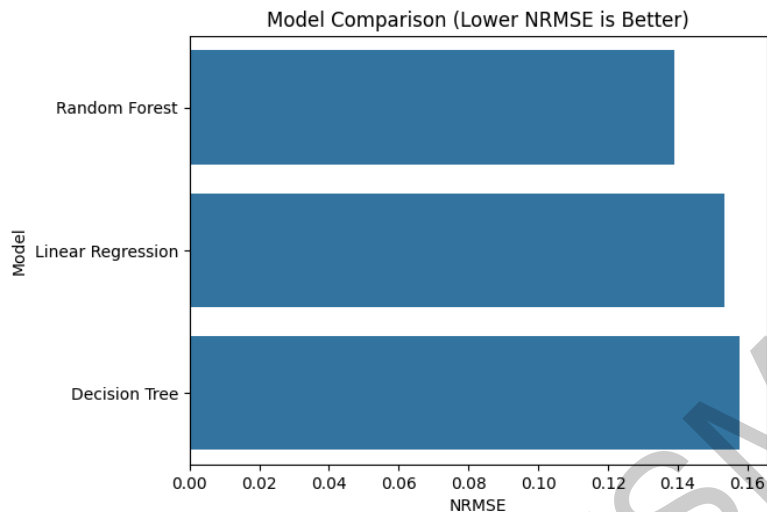
Keputusan merangkumi ketepatan model terbaik, keberkesanan pelaksanaan fungsi sistem dan tahap kebolegunaan antara muka pengguna. Hasil pengujian ini memberi gambaran sejauh mana sistem memenuhi keperluan fungsi yang telah ditetapkan serta memastikan prestasinya berada pada tahap yang memuaskan untuk digunakan secara praktikal.

Jadual 1.0 Output perbandingan model (model terbaik)

Model	RMSE	R2 Score	NRMSE
Random Forest	30.972885	0.970592	0.139022
Linear Regression	34.152684	0.964243	0.153295
Decision Tree	35.136127	0.962155	0.157709

Model Terbaik berdasarkan NRMSE: Random Forest

Rajah 1.0 pula menunjukkan hasil output pengiraan ketiga-tiga metrik bagi setiap model. Dapat dilihat bahawa model *Random Forest* mencatatkan nilai RMSE terendah (30.97) dan nilai NRMSE terendah (0.139), menunjukkan ketepatan ramalan yang lebih baik berbanding dua model lain. Walaupun ketiga-tiga model menunjukkan nilai R^2 melebihi 0.96, *Random Forest* tetap unggul dengan skor tertinggi (0.9706).



Rajah 1.1 Visualisasi perbandingan model terbaik

Rajah 1.1 menunjukkan visualisasi perbandingan nilai NRMSE antara model-model tersebut dalam bentuk carta bar. Visual ini memudahkan perbandingan secara terus dan menunjukkan dengan jelas bahawa *Random Forest* menghasilkan prestasi yang paling baik berdasarkan ketepatan ramalan.

Melalui ketiga-tiga ujian ini, dapat disimpulkan bahawa model *Random Forest Regressor* adalah model yang paling sesuai dan terbaik digunakan dalam sistem cadangan makanan ini. Ia memberikan keseimbangan terbaik antara ketepatan (nilai ralat rendah) dan penyesuaian terhadap data sebenar (R^2 tinggi), sekaligus dipilih sebagai model akhir yang digunakan dalam sistem.



Rajah 1.2 Hasil pengujian masa sistem

Rajah 1.2 menunjukkan hasil pengujian menunjukkan bahawa sistem berjaya menjana cadangan makanan dalam masa 0.07 saat, yang merupakan masa pemprosesan

yang sangat pantas dan memenuhi keperluan bagi penggunaan secara masa nyata (*real-time*). Ini membuktikan bahawa sistem ini efisien dan bersesuaian untuk digunakan oleh pengguna akhir tanpa mengalami kelewatan atau gangguan dari segi prestasi.

Food Recommendation System

Enter your personal data and get food suggestions that match your daily calorie needs (7 items).

Gender
☐ male ☒ female

Weight (kg)
 30 150 (current: 46)

Height (cm)
 120 210 (current: 143)

Age
 10 80 (current: 18)

Activity Level
☐ low ☐ moderate ☒ high

Buttons: Clear, Submit

Your personalized recommendation!

Rajah 1.3 Antara muka tanpa sebarang input pengguna

Rajah 1.3 menunjukkan skrin utama sistem sebelum sebarang data dimasukkan oleh pengguna. Pengguna akan melihat ruangan untuk memasukkan maklumat seperti umur, jantina, berat badan, tinggi, dan tahap aktiviti fizikal. Pada tahap ini, butang “Submit” disediakan untuk memulakan proses pengiraan keperluan kalori harian berdasarkan maklumat yang akan dimasukkan.

Food Recommendation System

Enter your personal data and get food suggestions that match your daily calorie needs (7 items).

Gender
☐ male ☒ female

Weight (kg)
 30 150 (current: 55)

Height (cm)
 120 210 (current: 153)

Age
 10 80 (current: 22)

Activity Level
☐ low ☒ moderate ☐ high

Buttons: Clear, Submit

Your personalized recommendation!

Estimated Daily Calories: 1915 kcal

Recommended Food Combinations (7 items):

1. Pita bread (274 kcal) + Pita (274 kcal) + Raisin bread (274 kcal) + Taquitos (274 kcal) + Sugar (274 kcal) + Pepperoni Pizza (273 kcal) + Grenadine (274 kcal)
Total: 1916 kcal
2. Pita bread (274 kcal) + Pita (274 kcal) + Raisin bread (274 kcal) + Taquitos (274 kcal) + Sugar (274 kcal) + Pepperoni Pizza (273 kcal) + French rolls (273 kcal)
Total: 1914 kcal
3. Pita bread (274 kcal) + Pita (274 kcal) + Raisin bread (274 kcal) + Taquitos (274 kcal) + Sugar (274 kcal) + Pepperoni Pizza (273 kcal) + Fruit preserves (274 kcal)
Total: 1916 kcal

Generated in 0.01 seconds

Rajah 1.4 Antara muka peringkat akhir selepas keputusan dipaparkan

Rajah 1.4 ini memaparkan keadaan sistem selepas pengguna memasukkan semua

maklumat diri yang diperlukan. Data seperti umur, jantina, berat badan, tinggi, dan tahap aktiviti fizikal telah dilengkapkan, menandakan sistem sedia untuk mengira keperluan kalori harian. Paparan ini membolehkan pengguna menyemak semula maklumat yang dimasukkan. Dalam rajah ini juga sudah ditunjukkan keputusan yang dihasilkan selepas pengguna menekan butang “Submit”. Sistem akan memaparkan jumlah keperluan kalori harian pengguna berserta senarai cadangan makanan yang mempunyai nilai kalori hampir dengan keperluan tersebut. Setiap cadangan makanan turut disertakan dengan maklumat nutrisi penting, membolehkan pengguna membuat pilihan makanan yang bijak dan sesuai dengan keperluan mereka.

Cadangan Penambahbaikan

Beberapa penambahbaikan boleh dilakukan seperti penambahan kategori diet (contohnya bebas gluten, rendah karbohidrat), integrasi dengan pangkalan data makanan tempatan/global melalui API, serta reka bentuk antara muka yang lebih interaktif termasuk paparan imej makanan dan pengelasan mengikut waktu makan. Fungsi pemantauan pengambilan makanan harian juga boleh dimasukkan sebagai ciri tambahan pada masa hadapan.

KESIMPULAN

Secara keseluruhan, projek ini telah berjaya membangunkan sistem pencadangan makanan diperibadikan menggunakan teknik Content-Based Filtering (CBF). Sistem ini memberikan saranan makanan berdasarkan keperluan kalori harian individu yang dikira menggunakan formula Harris-Benedict, dengan mengambil kira faktor seperti umur, jantina, tinggi, berat badan dan tahap aktiviti fizikal. Data yang digunakan terdiri daripada 1174 item makanan dengan 59 ciri nutrien, yang telah diproses dan dibersihkan bagi memastikan ketepatan cadangan.

Kekuatan Sistem

Kekuatan utama sistem ini ialah keupayaannya untuk memberikan cadangan makanan yang diperibadikan secara cepat dan tepat. Model Random Forest Regressor telah dipilih sebagai model utama kerana prestasinya yang terbaik dalam ujian metrik seperti RMSE, NRMSE dan R^2 . Penggunaan platform Gradio juga menjadikan antara muka sistem mesra pengguna, membolehkan interaksi mudah tanpa memerlukan log masuk, sesuai digunakan oleh pelbagai lapisan masyarakat.

Kelemahan Sistem

Namun begitu, sistem ini masih mempunyai beberapa kekangan seperti keterbatasan data makanan yang hanya bergantung kepada dataset statik. Sistem ini juga belum mengambil kira faktor penting lain seperti alahan makanan, diet khas, dan keperluan kesihatan tertentu. Selain itu, tiada integrasi dengan API makanan luaran yang boleh memperkayakan lagi cadangan makanan yang diberikan.

PENGHARGAAN

Penulis kajian ini ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Dr. Hazura binti Mohamed, penyelia penulis kajian ini yang telah memberi tunjuk ajar serta bimbingan untuk menyiapkan projek ini dengan jayanya.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

RUJUKAN

Wikipedia contributors. (2025, June 3). Harris–Benedict equation. Wikipedia.
https://en.wikipedia.org/wiki/Harris%E2%80%93Benedict_equation

Koul, N. (2024, November 28). Getting started with Gradio Python Library - Dr. Nimrita Koul - Medium.
<https://medium.com/@nimritakoul01/getting-started-with-gradio-python-library-49e59e363c66>

Jacob Murel, I. (2024, December 19). Content-based filtering. Content-Based Filtering.
https://www.ibm.com/think/topics/content-based-filtering?utm_source=

Nurin Jazlina binti Feroz (A195645)
Dr. Hazura binti Mohamed
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia