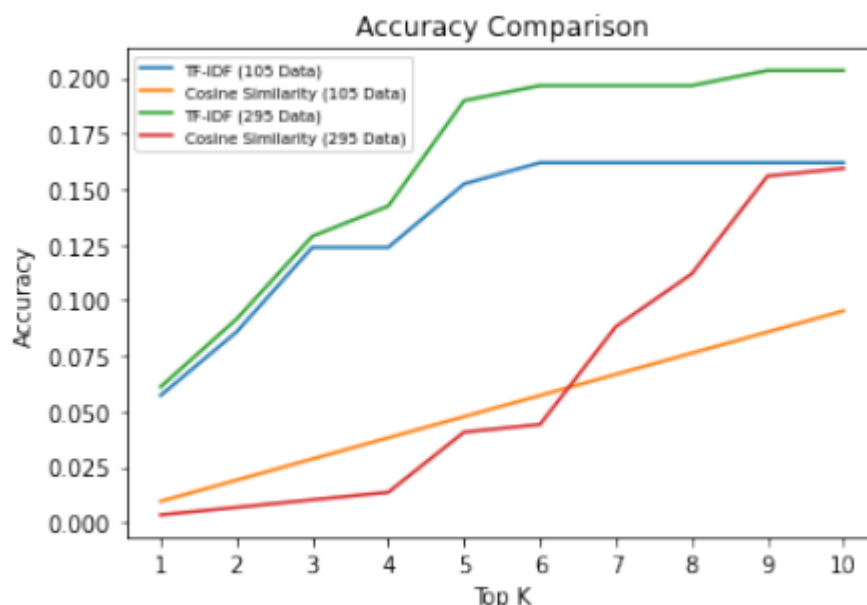


## Ujian Akhir Semester Kelas STBI

Nama: Hayyu Ilham Wicaksono

Nomer Mahasiswa: 483952

1. Metode kueri pencarian dan *similarity* yang digunakan yaitu menggunakan TF-IDF dan *cosine similarity*. Pencarian kemiripan dengan menggabungkan dataset Rekmed dengan dataset dari repositori GitHub *Disease Detection based on Symptoms with treatment recommendation* untuk mendapatkan data yang lebih baik. Pencarian kemiripan berdasarkan gejala – gejala pasien dengan diagnosis yang tersedia pada dataset.
2. Evaluasi sistem menggunakan *Top k-accuracy-score*. Model evaluasi ini untuk menggeneralisasi prediksi skor suatu sistem rekomendasi tanpa proses training data. Evaluasi ini kami gunakan karena data evaluasi yang kami gunakan adalah data uji, dan tidak menggunakan data training. Berdasarkan model evaluasi tersebut kami menguji dua daftar gejala, yaitu 105 daftar gejala dan label penyakitnya, sebelum dilakukan augmentasi data. Kemudian 190 data yang telah dilakukan augmentasi dataset. Hasilnya menunjukkan pada metode TF-IDF tanpa melakukan augmentasi dengan  $k=10$  akurasi 16%, sedangkan setelah melakukan augmentasi meningkat 4% menjadi 20%. Peningkatan hasil akurasi pada data uji yang telah diaugmentasi ini dianalisis bahwa masukkan gejala yang relevan tentunya akan meningkatkan akurasi prediksi penyakitnya. Pada data uji yang kami gunakan pada penelitian ini meskipun data telah diaugmentasi menjadi hampir 3 kali lebih besar dari data uji sebelumnya diperoleh peningkatan hasil akurasi yang tidak terlalu besar. Hal tersebut dianggap bahwa masukkan gejala yang tidak relevan terhadap penyakitnya tidak akan mempengaruhi hasil prediksinya. Berikut merupakan grafik evaluasi sistem prediksi penyakit ini.



3. Sistem rekomendasi merupakan sistem yang dibangun untuk para user atau pengguna agar mendapatkan produk atau item atau data informasi yang diinginkan dan rekomendasi tertentu. Output rekomendasi berdasarkan sumber informasi yang telah diurutkan sesuai dengan dataset. Sedangkan sistem tanya jawab berfokus pada jawaban yang tersedia pada koleksi dan hanya menampilkan informasi yang ada pada koleksi. Sistem yang kami termasuk

dalam sistem rekomendasi karena sistem yang dibangun tidak hanya memberikan informasi terkait gejala-gejala yang mirip dengan koleksi dalam dataset, namun juga memberikan rekomendasi beberapa diagnosis yang paling relevan dan detail terkait penyakit tertentu yang kemungkinan diderita oleh pasien.

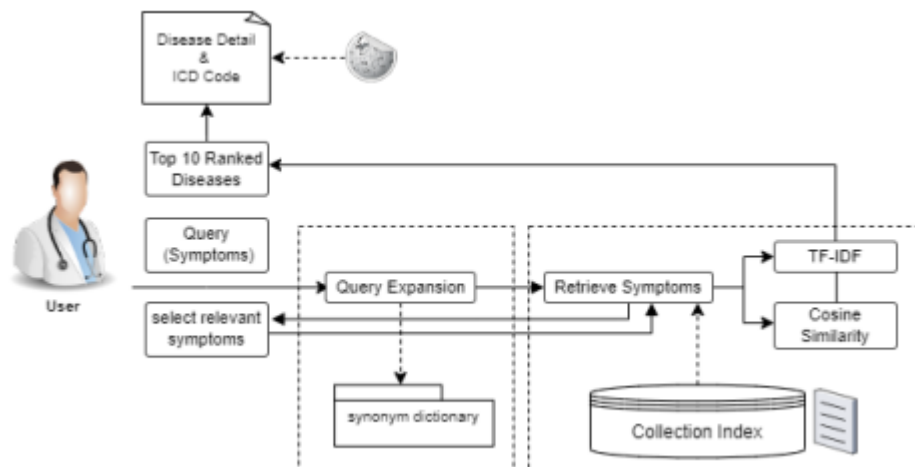
#### 4. Rangkuman:

##### A. Introduction

Teknologi pencarian diagnosis penyakit digunakan untuk menemukan informasi penyakit yang sesuai dan relevan secara cepat. Klasifikasi suatu gejala penyakit ke dalam kode ICD, untuk membantu memberikan rekomendasi diagnosis penyakit sangat melelahkan dan memakan waktu. Seorang profesional yang telah memiliki banyak pengalaman membutuhkan waktu sekitar 20 menit per kasus. Sehingga, pengembangan sistem untuk meningkatkan akurasi diagnosis dan mengurangi waktu sangat diperlukan dalam menentukan diagnosis pasien. Kami mengusulkan sistem untuk mendiagnosis primordial suatu penyakit yang tertera pada ICD 10 berdasarkan gejala yang dialami. Untuk meningkatkan akurasi prediksi, peneliti melakukan augmentasi dataset agar kemungkinan variasi gejala yang mirip dapat dikenali optimal. Evaluasi dilakukan dengan metode *Top k-accuracy*, dengan memasukkan seluruh gejala penyakit pada dataset.

##### B. Methods

Sistem yang kami usulkan pada penelitian ini memiliki desain sebagai berikut.



Berdasarkan rancangan desain tersebut kami melakukan proses *preprocessing* terhadap dataset, meliputi *casefolding*, tokenisasi, dan *stemming*. Hasil dari proses *preprocessing*, kami lakukan proses perluasan kueri dengan sinonim dari masukkan gejala. Perluasan dilakukan dengan kombinasi dari daftar sinonim sehingga menghasilkan daftar gejala yang cocok yang telah dibubuhkan indeks sehingga pengguna dapat mendapatkan gejala yang sesuai dengan indeksnya. Kemudian melakukan prediksi dengan metode TF-IDF dan *cosine similarity*, berdasarkan model ini ditampilkan 10 prediksi diagnosis teratas yang paling sesuai. Output sistem tidak hanya prediksi diagnosis penyakit tetapi juga beserta detail penyakit tertentu yang pasien pilih, kami mengambil detail penyakit berdasarkan dataset infobox Wikipedia.

##### C. Discussion

Hasil penelitian ini menunjukkan bahwa pada TF-IDF memiliki akurasi prediksi penyakit dengan k prediksi dengan nilai tertinggi yaitu pada k = 10 akurasi hanya 16% sedangkan setelah augmentasi data ujinya, akurasi k = 10 meningkat dengan kenaikan pada k = 1 sekitar 1% dan k = 10 sekitar 4% menjadi 20%. Hasil akurasi pada metode *cosine similarity* secara

keseluruhan lebih rendah, prediksi tertinggi pada  $k = 10$  yaitu 9,5% dan dengan data uji yang diaugmentasi mengalami penurunan akurasi untuk  $k = 1$  sampai  $k = 6$ . Nilai akurasi meningkat pada  $k = 7$  hingga  $k = 10$ , dengan akurasi paling tinggi mendekati 16%. Dari penelitian ini, hasil uji dengan proses augmentasi dataset secara umum menghasilkan akurasi yang lebih baik. Selain itu, hasil prediksi menggunakan metode TF-IDF lebih baik dari pada menggunakan metode cosine similarity. Berikut merupakan gambar hasil evaluasi dalam bentuk table dari sistem yang kami kembangkan.

**PERBANDINGAN EVALUASI TOP  $k$  AKURASI  
PADA TF-IDF DAN *Cosine Similarity***

$k$	TF-IDF (105 Data)	Cosine Similarity (105 Data)	TF-IDF (295 Data)	Cosine Similarity (295 Data)
1	0.057143	0.009524	0.061017	0.003390
2	0.085714	0.019048	0.091525	0.006780
3	0.123810	0.028571	0.128814	0.010169
4	0.123810	0.038095	0.142373	0.013559
5	0.152381	0.047619	0.189831	0.040678
6	0.161905	0.057143	0.196610	0.044068
7	0.161905	0.066667	0.196610	0.088136
8	0.161905	0.076190	0.196610	0.111864
9	0.161905	0.085714	0.203390	0.155932
10	0.161905	0.095238	0.203390	0.159322

#### D. References

M. Mustakim and R. Wardoyo, "Survey model-model pencarian informasi rekam medik elektronik," JISKA (Jurnal Informatika Sunan Kalijaga), vol. 3, no. 3, p. 132–144, Aug. 2019. [Online]. Available: <https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/33-01>

R. Silalahi and E. Sinaga, "Perencanaan implementasi rekam medis elektronik dalam pengelolaan unit rekam medis klinik pratama romana," Jurnal Manajemen Informasi Kesehatan Indonesia, vol. 7, p. 22, 03 2019.

Commonwealth of Australia, "MBS Telehealth Services from 1 July 2022," 2022. [Online]. Available: <http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Factsheet-telehealth-1July22>

V. K and S. Jyothi, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," International Journal of Computer Applications, vol. 19, 04 2011.

C. D. Manning, P. Raghavan, and H. Schutze, "Introduction to Information Retrieval. USA: Cambridge University Press, 2008.

A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," in 2016 4th International Conference on Cyber and IT Service Management, 2016, pp. 1–6.

K. Park, J. Hong, and W. Kim, "A methodology combining cosine similarity with classifier for text classification," *Applied Artificial Intelligence*, vol. 34, pp. 1–16, 02 2020.

scikit-learn developers, "Metrics and scoring: quantifying the quality of predictions," 2022. [Online]. Available: [https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html)

Penjelasan terkait *multimedia information retrieval system*. Sistem yang kami kembangkan termasuk *multimedia information retrieval* karena tidak hanya mengolah data teks, namun sistem ini berbasis *content base information retrieval*, yaitu database yang kami ekstrak, kemudian dilakukan *preprocessing* dan dideskripsikan oleh vector fitur. Vektor fitur tersebut kemudian dihitung dan melakukan *ranking*. *Content base information retrieval* termasuk dalam model *multimedia information retrieval*.