

APLIKASI CV MATCHER UNTUK MELIHAT KECOCOKAN DAFTAR RIWAYAT HIDUP DENGAN LOWONGAN PEKERJAAN MENGGUNAKAN MACHINE LEARNING DAN METODE COSINE SIMILARITY BERBASIS WEB

Wildan Adhitya Geraldine, Yudi Herdiana, Yusuf Muharam

ABSTRAK: Tugas akhir skripsi merupakan salah satu syarat dalam menyelesaikan program strata I (satu). Sebelum memulai pengerjaan skripsi ini, dilakukan terlebih dahulu pengajuan proposal skripsi, yang mana akan menjadi sebuah dasar dalam pengerjaan tugas akhir skripsi. Berdasarkan laporan Badan Pusat Statistik (BPS), jumlah pengangguran di Indonesia pada Agustus 2021 adalah sebesar 9,10 juta penduduk. Jumlah itu menurun dibanding jumlah pengangguran setahun sebelumnya yang mencapai 9,77 juta orang. Permasalahan yang saat ini banyak dijumpai adalah ketatnya persaingan dalam memperoleh pekerjaan sehingga para pelamar kerja harus mempersiapkan lamaran kerja dengan matang, salah satunya yaitu berkas CV yang dapat menunjukkan keunggulan pelamar kerja. Tujuan dari penelitian ini adalah membuat aplikasi *CV Matcher* yang dapat membantu pelamar kerja dalam mengecek kesesuaian *Curriculum Vitae* yang mereka buat dengan kriteria kandidat yang sudah ditentukan oleh perusahaan yang dituju. Penelitian ini menggunakan metode kualitatif yaitu dengan melakukan studi literatur. Selanjutnya dilanjutkan dengan tahapan perancangan sebagai gambaran umum dari aplikasi yang dibangun. Pada penelitian ini, model pengembangan pada perancangan sistem yang digunakan oleh penulis adalah model *AI Project Cycle*. Ada 6 tahapan pada model *AI Project Cycle* diantaranya adalah *Problem Scoping*, *Data Acquisition*, *Data Exploration*, *Modelling*, *Evaluation*, *Deployment*. Kemudian secara detail mengenai rancangan aplikasi pada tahap *Deployment AI Project Cycle* yaitu dengan menggunakan *UML (Unified Modelling Language)* yang meliputi *use case diagram*, dan *activity diagram*. Kemudian untuk beberapa *Software* yang digunakan dalam melakukan penelitian ini adalah *Balsamiq Mockup* untuk merancang desain tampilan aplikasi. *Google Colaboratory* digunakan untuk merancang model *Machine Learning* yang akan digunakan. Selanjutnya untuk bahasa pemrograman yang digunakan adalah python 2.7. Untuk perancangan aplikasi menggunakan *Unified Model Language (UML)* dan *HTML* serta *CSS Bootstrap* untuk implementasi tampilannya. Hasil dari penelitian ini adalah dengan adanya aplikasi *CV Matcher* yang dapat mencocokkan CV pelamar kerja dengan deskripsi lowongan pekerjaan yang akan dilamar.

Kata Kunci: *Artificial Intelligence*, CV, Lowongan Pekerjaan, Pelamar Kerja

ABSTRACT: *The final thesis is one of the requirements in completing the strata I (one) program. Before starting to work on this thesis, it is necessary to submit a thesis proposal first, which will be the basis for working on the final thesis. Based on a report from the Central Statistics Agency (BPS), the number of unemployed in Indonesia in August 2021 was 9.10 million people. That number decreased compared to the number of unemployed a year earlier which reached 9.77 million people. The problem that is currently encountered is the intense competition in getting a job, so job applicants must prepare job applications carefully, one of which is a CV file that can show the advantages of job applicants. The purpose of this study is to Create a CV Matcher application that can assist job applicants in checking the suitability of the Curriculum Vitae they have made with the candidate criteria that have been determined by the intended company. This study uses a qualitative method, namely by conducting a literature study. Then proceed with the design stage as a general description of the application being built. In this study, the development model in system design used by the author is the AI Project Cycle model. There are 6 stages in the AI Project Cycle model including Problem Scoping, Data Acquisition, Data Exploration, Modelling, Evaluation, Deployment. Then in*

detail about the application design at the AI Project Cycle Deployment stage, namely by using UML (Unified Modelling Language) which includes Use case diagrams, and activity diagrams. Then for some of the Software used in doing this design is the Balsamiq Mockup to design the display design of the application. Google Collaboratory is used to design the Machine Learning model that will be used. Next for the programming language used is python 2.7. For application design using Unified Model Language (UML) and HTML and CSS Bootstrap for the implementation of the appearance. The result of this research is that there is a CV Matcher application that can match the CV of job applicants with a description of the job vacancies to be applied for.

Keywords: Application, Certificate, Cover Letter, Application for Leave, Website

PENDAHULUAN

Berdasarkan laporan Badan Pusat Statistik (BPS), jumlah pengangguran di Indonesia pada Agustus 2021 adalah sebesar 9,10 juta penduduk. Jumlah itu menurun dibanding jumlah pengangguran setahun sebelumnya yang mencapai 9,77 juta orang (BPS: *Tingkat Pengangguran Terbuka Pada Agustus 2021 Turun 0,58 Persen Dibanding Agustus 2020 - Data Tempo.Co*, 2021). Hal ini menunjukkan bahwa masih banyak nya jumlah pengangguran. Hal ini tentunya menjadi salah satu faktor kegagalan para pelamar kerja dalam proses seleksi penerimaan karyawan di suatu perusahaan. Livecareer.com merupakan sebuah website penyedia informasi mengenai lowongan pekerjaan yang mencakup dari beberapa perusahaan di seluruh dunia seperti *Amazon*, *Google*, *AWS*, dan perusahaan lainnya. Selain itu livecareer.com juga merupakan website penyedia informasi tentang pembuatan daftar riwayat hidup bagi para pelamar kerja. Livecareer.com juga menyediakan berbagai artikel informasi yang dapat berguna bagi para pelamar kerja untuk melamar suatu pekerjaan. Hal ini juga didukung oleh berbagai fitur yang disediakan seperti fitur pembuatan daftar riwayat hidup dengan ribuan desain menarik yang bisa digunakan.

Disamping berbagai fitur menarik yang telah disediakan oleh website livecareer.com, namun dalam penelusuran penulis pada website tersebut masih ada beberapa fitur yang kurang bagi pengguna terutama pelamar kerja untuk menyiapkan daftar riwayat hidup atau berkas lamaran yang dimilikinya. Salah satunya adalah penulis menemukan bahwa pada website livecareer.com masih belum memiliki fitur untuk melakukan pengecekan kecocokan berkas lamaran pelamar kerja dengan lowongan pekerjaan yang akan dilamar. Sehingga hal tersebut dirasa kurang meyakinkan pelamar kerja untuk memiliki peluang diterima terhadap suatu lowongan pekerjaan tersebut. Berdasarkan jurnal yang di terbitkan oleh Jurnal Aghinya Stiesnu, (Handoko, 2020) menyatakan bahwa rekrutmen adalah proses mencari, menemukan dan menarik para pelamar yang kapabel untuk dipekerjakan dalam suatu organisasi. Selain itu, rekrutmen juga dapat diartikan sebagai proses untuk mendapatkan sejumlah SDM atau karyawan yang berkualitas untuk menduduki suatu jabatan atau pekerjaan disebuah perusahaan. Pada sebuah proses rekrutmen, akan diadakan tahap seleksi. Seleksi adalah proses memilih pelamar sampai dengan memutuskan pelamar mana yang diterima atau yang ditolak untuk dijadikan pegawai. Salah satu tahapan

seleksi pada lowongan kerja adalah tahapan seleksi surat - surat lamaran. Maka dari itu, berkas lamaran seperti daftar riwayat hidup atau surat lamaran harus disiapkan sebaik-baiknya oleh pelamar kerja, agar dapat sesuai dengan lowongan pekerjaan yang akan dilamar. Maka dari itu, penulis mengangkat topik penelitian ini dengan judul "Aplikasi CV *Matcher* Untuk Melihat Kecocokan Daftar Riwayat Hidup Dengan Lowongan Pekerjaan Menggunakan *Machine Learning* Dan Metode *Cosine Similarity* Berbasis Web". Sehingga diharapkan aplikasi ini dapat ikut membantu para pelamar kerja untuk mengetahui kecocokan daftar riwayat hidup yang dimiliki dengan deskripsi lowongan pekerjaan yang akan dilamarnya.

METODA

Dalam penelitian ini, penulis menggunakan metode kualitatif yaitu dengan melakukan studi literatur. Selanjutnya dilanjutkan dengan tahapan perancangan sebagai gambaran umum dari aplikasi yang dibangun. Pada tahap ini, model pengembangan pada perancangan sistem yang digunakan oleh penulis adalah model *AI Project Cycle*. Ada 6 tahapan pada model *AI Project Cycle* diantaranya adalah *Problem Scoping*, *Data Acquisition*, *Data Exploration*, *Modelling*, *Evaluation*, *Deployment*. Kemudian secara detail mengenai rancangan aplikasi pada tahap *Deployment AI Project Cycle* yaitu dengan menggunakan UML (*Unified Modelling Language*) yang meliputi *use case diagram*, dan *activity diagram*.

HASIL DAN PEMBAHASAN

Menurut Arthur Samuel *Machine Learning* adalah sebuah pertanyaan "how can computers learn to solve problems without being explicitly

programmed?" yaitu bagaimana agar komputer dapat berjalan untuk memecahkan masalah sendiri tanpa harus diprogram secara eksplisit (Samuel, 1988). Sedangkan menurut Mehryar Mohri "*Machine Learning can be broadly defined as computational methods using experience to improve Performance or to make accurate predictions.*" Bersama kawan-kawannya Mohri mendefinisikan pengertian *Machine Learning* sebagai metode komputasi yang memanfaatkan experience untuk meningkatkan akurasi prediksi (Mehryar Mohri -- *Foundations of Machine Learning - Book*, 2022). Adapun menurut Ian Goodfellow dalam bukunya mendefinisikan sebagai berikut: *Machine Learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated Functions* (Deep Learning, 2022).

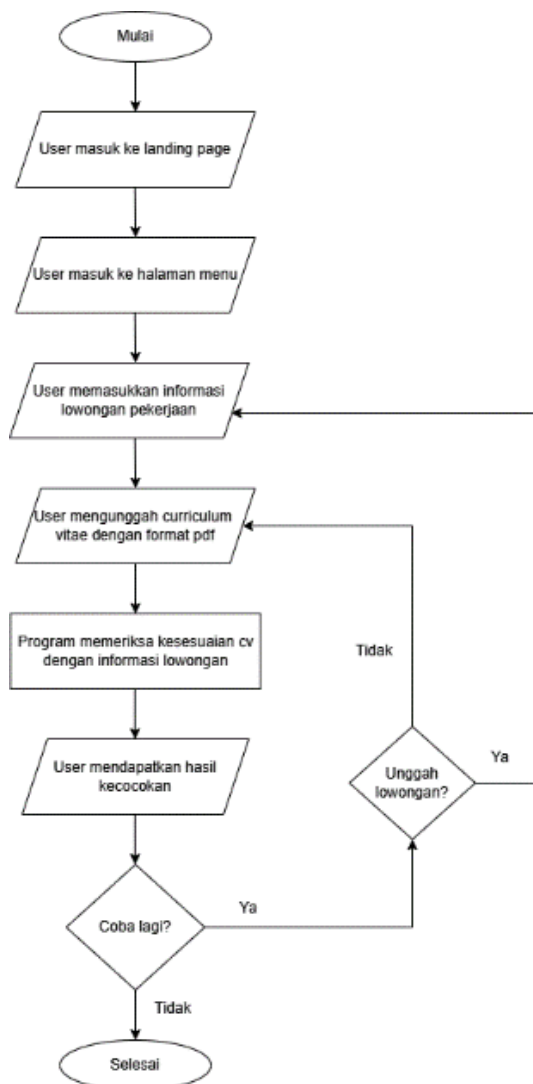
Natural Processing Language (NLP) adalah cabang ilmu komputer, linguistik, dan kecerdasan buatan yang mengkaji interaksi antara komputer dan bahasa (alami) manusia, khususnya cara memprogram komputer untuk mengolah data bahasa alami dalam jumlah besar. Hasilnya adalah komputer mampu "memahami" isi dokumen, termasuk nuansa bahasa di dalamnya. Dengan ini, komputer dapat dengan akurat mengambil informasi dan wawasan dari dokumen sekaligus mengelompokkan dan menata dokumen-dokumen itu sendiri (*Pengolahan Bahasa Alami - Wikipedia Bahasa Indonesia, Ensiklopedia Bebas*, 2022).

Berdasarkan artikel dari hendroprasetyo.com, *Cosine Similarity* adalah "ukuran kesamaan", salah satu implementasinya adalah pada kasus mencari tingkat kemiripan teks pada teks itu sendiri atau *sentence/kalimat*. Kemiripan teks bisa

kita gunakan untuk membuat steganografi ataupun steganalisis linguistik.

AI Project Cycle, yaitu merupakan sebuah kerangka kerja yang dapat digunakan untuk melakukan development sebuah proyek *Artificial Intelligence* (Handbook, n.d.). Proses dari *AI Project Cycle* tersebut terdiri atas 6 tahap, yaitu sebagai berikut : *Problem Scoping, Data Acquisition, Data Exploration Modelling, Evaluation, Deploymen*.

1. Flowchart aplikasi yang akan dibangun



2. Perancangan Model yg akan digunakan

| No. | Model | Hyperparameter | Accuracy | Recall | Precision | F1-Score | Waktu Training | Waktu Prediksi |
|-----|--------------------------|---|----------|--------|-----------|----------|----------------|----------------|
| 1. | Logistic Regression | <code>solver='liblinear', penalty='l2', C=1, multi_class='auto', random_state=0</code> | 72.05% | 72.05% | 72.84% | 71.15 % | 5.29 s | 0.08 s |
| 2. | Decision tree Classifier | <code>criterion='gini', splitter='random', max_depth=100, random_state=0</code> | 66.84% | 66.84% | 67.84% | 66.68 % | 10.87 s | 0.03 s |
| 3. | Random Forest Classifier | <code>criterion='gini', n_estimators=100, max_depth=150, random_state=0</code> | 79.69% | 79.69% | 80.74% | 79.01 % | 10.4 s | 0.13 s |
| 4. | Multinomial Naive Bayes | <code>alpha=0.01</code> | 67.36% | 67.36% | 68.18% | 66.75 % | 0.34 s | 0.05 s |
| 5. | KNN Classifier | <code>n_neighbors=5, weights='distance', p=2, metric='minkowski'</code> | 41.32% | 41.32% | 64.26% | 40.22 % | 0.04 s | 1.77 s |
| 6. | SVM | <code>kernel='linear', gamma='scale', C=10, decision_function_shape='ov', random_state=0</code> | 77.95% | 77.95% | 78.87% | 77.97 % | 174.03 s | 27.51 s |

Berdasarkan tabel di atas, maka model yg akan digunakan adalah model *Random Forest Classifier*. Model tersebut dinilai karena memiliki nilai *Accuracy, Recall, Precision*, dan *F1-Score* yg lebih baik diantara model lainnya.

3. Pengujian Sistem

Berikut adalah daftar pengujian yang dilakukan pada system aplikasi tersebut secara fungsionalitas tanpa menguji kode programnya.

| No | Item Uji | Skenario Pengujian | Hasil yang diharapkan | Hasil pengujian | Kesimpulan |
|----|-----------------|--------------------|------------------------|-----------------|------------|
| 1 | Halaman Beranda | Masuk ke aplikasi | Pengguna dapat melihat | Sesuai harapan | Valid |

| | | | | | |
|---|--------------------------|---|---|----------------|-------|
| | | | tampilan awal aplikasi | | |
| 2 | Halaman Pengembang | Klik menu halaman pengembang | Pengguna dapat melihat tampilan pengembang aplikasi | Sesuai harapan | Valid |
| 3 | Halaman Informasi | Klik halaman informasi | Pengguna dapat melihat tampilan informasi aplikasi | Sesuai harapan | Valid |
| 4 | Halaman Pengecekan | Mengunggah berkas CV dan Lowongan Pekerjaan | Pengguna dapat mengecek berkas CV dengan Lowongan Pekerjaan | Sesuai harapan | Valid |
| 5 | Halaman Hasil Pengecekan | Klik tombol Check pada halaman pengecekan | Pengguna dapat melihat hasil pengecekan berkas CV dengan Lowongan Pekerjaan | Sesuai harapan | Valid |

SIMPULAN

Berikut adalah simpulan dari proyek skripsi ini adalah sebagai berikut.

1. *CV Matcher* adalah aplikasi yang dapat digunakan oleh pelamar kerja untuk mengecek skor *Curriculum Vitae*, yakni apakah *Curriculum Vitae* yang telah dibuat sudah sesuai dengan kriteria pada deskripsi lowongan pekerjaan yang ditentukan oleh employer.
2. Pengembangan aplikasi *CV Matcher* dilaksanakan dengan pendekatan *AI Project Cycle*. Hal ini dimulai dari *Problem Scoping* dengan 4Ws; *Data Acquisition* dengan sumber data dari Kaggle; *Data Exploration* untuk memahami *dataset* serta mengecek

ketidakseimbangan data; *Modelling* untuk konversi *file CV* menjadi teks, klasifikasi kategori CV dengan model klasifikasi *multiclass*, dan scoring CV dengan *Cosine Similarity*; *Evaluation* untuk memilih *Random Forest Classifier* sebagai model klasifikasi *multiclass* terbaik; hingga *Deployment* untuk memasukkan model *Machine Learning* ke aplikasi berbasis website agar bisa digunakan.

3. Aplikasi *CV Matcher* ini dikembangkan dengan *Python* untuk pemodelan *Machine Learning*. Sementara itu, terkait *Deployment*, digunakan HTML, dan CSS *Bootstrap* sebagai *front-end*; *Flask* sebagai *back-end*; *tools* untuk repositori berupa *Github*; serta *tools* untuk hosting berupa Heroku versi 18.

PUSTAKA ACUAN

- Badan Pusat Statistik. (2022). Retrieved August 20, 2022, from <https://www.bps.go.id/pressrelease/2021/11/05/1816/agustus-2021-tingkat-pengangguran-terbuka--tpt--sebesar-6-49-persen.html>
- Burges, C. J. C., & Burges, C. J. C. (1998). A Tutorial on *Support Vector Machines* for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121–167. <https://www.microsoft.com/en-us/research/publication/a-tutorial-on-support-vector-Machines-for-pattern-recognition/>
- Chang, C. C., & Lin, C. J. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3). <https://doi.org/10.1145/1961189.1961199>
- Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to *Support Vector Machines* and Other Kernel-based Learning Methods.

- An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. <https://doi.org/10.1017/CBO9780511801389>
- Deep Learning. (2022). Retrieved August 20, 2022, from <https://www.deeplearningbook.org/>
- Deshpande, A. R., & L.M.R.J, L. (2015). *Text Summarization using Clustering Technique and SVM Technique*. *International Journal of applied Engineering Research*, 10(10), 25511–25519.
- Gunawan, K. I., & Santoso, J. (2021). *Multilabel Text Classification Menggunakan SVM dan Doc2Vec Classification Pada Dokumen Berita Bahasa Indonesia*. *Journal of Information System, Graphics, Hospitality and Technology*, 3(01), 29–38. <https://doi.org/10.37823/insight.v3i01.126>
- Handbook, F. (n.d.). *CLASS 10 Curated with support from Intel® AcKnowledgements*.
- Handoko. (2020). Analisis Proses Rekrutmen, Seleksi dan Penempatan Karyawan di PT MRT Jakarta. *Jurnal Aghniya*, 3(1), 137–138. <https://ejournal.stiesnu-bengkulu.ac.id/index.php/aghniya/article/view/41>
- Hardiyanto, H. (Hardiyanto), Abdussomad, A. (Abdussomad), Haryadi, E. (Eko), Sopandi, R. (Robi), & Asep, A. (Asep). (2019). Penerapan Model Waterfall dan Uml dalam Rancang Bangun Program Pembelian Barangberorientasi Objek pada PT. FUJITA INDONESIA. *Jurnal Interkom*, 13(4), 4–11. <https://doi.org/10.35969/INTERKO M.V13I4.37>
- Ladwani, V. M. (2018). *Support Vector Machines and applications. Computer Vision: Concepts, Methodologies, Tools, and applications, February, 1381–1390*. <https://doi.org/10.4018/978-1-5225-5204-8.ch057>
- Mehryar Mohri -- *Foundations of Machine Learning - Book*. (n.d.). Retrieved August 20, 2022, from <https://cs.nyu.edu/~mohri/mlbook/>
- Natural language Processing*. (2022). Retrieved August 20, 2022, from <https://socs.binus.ac.id/2013/06/22/natural-language-Processing/>
- Octaviani, P. A., Wilandari, Y., & Ispriyanti, D. (2014). PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG. *Undefined*.
- Pengolahan bahasa alami - Wikipedia bahasa Indonesia, ensiklopedia bebas*. (2022). Retrieved August 20, 2022, from https://id.wikipedia.org/wiki/Pengolahan_bahasa_alami
- Perbedaan Cosine Similarity dan Cosine Distance - Hendro Prasetyo*. (2022). Retrieved August 20, 2022, from <https://hendroprasetyo.com/perbedaan-Cosine-Similarity-dan-Cosine-distance/#.YwBbwnZBzDd>
- Samuel, A. L. (1988). Some Studies in *Machine Learning Using the Game of Checkers*. II—Recent Progress. *Computer Games I*, 366–400. https://doi.org/10.1007/978-1-4613-8716-9_15
- Wahid, D. H., & SN, A. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan *Hybrid TF-IDF dan Cosine Similarity*. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), 207–218. <https://doi.org/10.22146/IJCCS.16625>
- Waruwu, T. S., & Nasution, S. (2018).

Pengembangan Keamanan Web Login Portal Dosen Menggunakan Unified *Modelling* Language (UML). *Jurnal Mahajana Informasi*, 3(1), 34–40.

Willy, W., Rini, D. P., & Samsuryadi, S. (2021). Perbandingan Algoritma *Random Forest Classifier*, *Support Vector Machine* dan *Logistic Regression Classifier* Pada Masalah *High Dimension* (Studi Kasus: Klasifikasi *Fake News*). *Jurnal Media Informatika Budidarma*, 5(4), 1720. <https://doi.org/10.30865/mib.v5i4.3177>