



Penerapan Algoritma Logistic Regression untuk Memprediksi Keberhasilan Terapi Kutil (Cryotherapy)

Application of Logistic Regression Algorithm to Predict the Success of Wart Therapy (Cryotherapy)

**Muhammad Fiqih Ainurohman^{1*}, Hasbi Firmansyah², Wahyu Asriyani³,
Rizki Prasetyo Tulodo⁴**

Universitas Pancasakti Tegal

Email: kenriyan2004@gmail.com^{1*}, hasbifirmansyah@upstegal.ac.id², asriyani1409@gmail.com³
Rizki.prasetyo.tulodo@gmail.com⁴

Article Info

Article history :

Received : 19-12-2025

Revised : 20-12-2025

Accepted : 22-12-2025

Pulished : 25-12-2025

Abstract

Cryotherapy, which is a type of freezing treatment, is often used to treat warts, but how well it works can be different for each person. This study tries to create a model that can predict whether Cryotherapy will be successful using Data Mining and the Logistic Regression method. The data comes from the UCI Machine Learning Repository and includes information from 90 patients with 6 different clinical factors: Age, Gender, Therapy Time, Number of Warts, Type, and Area. The study was done using RapidMiner software and tested with a 10-Fold Cross Validation method. The results show that the Logistic Regression model can predict the success of the treatment with an accuracy of 86.67%. The study also found that the age of young patients between 15 and 20 years old has the biggest impact on how well the treatment works.

Keywords: Data Mining, Logistic Regression, Cryotherapy

Abstrak

Cryotherapy (terapi beku) adalah cara umum untuk mengobati kutil, tetapi tingkat keberhasilannya berbeda-beda antar pasien. Penelitian ini bertujuan membuat model untuk memprediksi apakah Cryotherapy berhasil atau tidak, menggunakan teknik Data Mining dengan algoritma Logistic Regression. Data yang digunakan berasal dari UCI Machine Learning Repository, yang berisi 90 data pasien dan 6 informasi klinis, yaitu usia, jenis kelamin, durasi terapi, jumlah kutil, jenis kutil, dan luas area kutil. Penelitian dilakukan dengan menggunakan perangkat lunak RapidMiner dan metode validasi 10-Fold Cross Validation. Hasil menunjukkan algoritma Regresi Logistik mampu memprediksi keberhasilan terapi dengan akurasi sebesar 86,67%. Penelitian juga menemukan bahwa variabel usia pasien yang muda, yaitu antara 15 sampai 20 tahun, berpengaruh paling besar terhadap hasil pengobatan.

Kata Kunci : Data Mining, Regresi Logistik, Cryotherapy

PENDAHULUAN

Kesehatan kulit sangat penting untuk menjaga kualitas hidup manusia. Salah satu masalah kulit yang sering terjadi di seluruh dunia adalah kutil, atau disebut juga verruca vulgaris. Penyakit ini terjadi karena infeksi virus Human Papillomavirus (HPV) yang menyerang sel kulit di lapisan terluar kulit, menyebabkan pertumbuhan sel yang berlebihan. Meskipun sering hanya dianggap sebagai masalah penampilan, kutil bisa membuat nyeri, membuat seseorang merasa tidak nyaman secara psikologis, bahkan mengganggu fungsi tubuh, terutama jika muncul di area seperti telapak tangan atau kaki. Karena kutil bisa menular, maka penting untuk segera diobati agar tidak menyebar ke orang lain.(Degree et al., 2012)



Dalam pengobatan kutil di bidang dermatologi, ada beberapa cara untuk mengatasi masalah tersebut, sepertioleskan obat asam salisilat, prosedur bedah listrik, terapi laser, serta Cryotherapy. Di antara metode ini, Cryotherapy atau terapi beku yang menggunakan nitrogen cair dengan suhu -196°C sering dipilih sebagai metode utama karena prosedurnya cepat, tidak memakan banyak waktu, dan tidak memerlukan operasi yang invasif. Cara kerjanya adalah dengan membekukan air di dalam sel kutil, sehingga menyebabkan rusaknya membran sel dan akhirnya menghancurkan jaringan yang terinfeksi virus tersebut.(Dreiseitl & Ohno-machado, 2003)

Namun, efektivitas terapi krioterapi tidak selalu pasti dan bisa berbeda-beda pada setiap orang. Berdasarkan penelitian klinis, tingkat kesuksesan pengobatan ini dipengaruhi oleh berbagai faktor, seperti usia pasien, lama infeksi, lokasi tumbuh kutil, jenis kutil, hingga respons tubuh pasien terhadap sistem imun. Ketidakpastian ini menimbulkan masalah besar, yaitu dokter kesulitan memperkirakan dengan tepat apakah pasien bisa sembuh dalam satu kali sesi atau perlu menjalani pengobatan berulang. Jika terapi pertama gagal, pasien sering harus menerima sesi tambahan yang bisa menyakitkan, meningkatkan risiko efek samping seperti kulit makin pucat atau luka parut, serta membuat pasien merasa kewalahan secara biaya dan waktu.(Verma et al., 2019)

Hingga kini, prediksi keberhasilan pengobatan biasanya masih bergantung pada perasaan dan pengalaman dokter secara pribadi, tanpa ada alat yang bisa mengukur secara objektif. Hal ini memberi kesempatan bagi penggunaan teknologi informasi, terutama bidang Medical Data Mining, untuk memberikan solusi yang lebih akurat. Data Mining memungkinkan menemukan pola-pola tersembunyi dari data rekam medis yang sudah ada, sehingga bisa dibuat model prediksi yang lebih cerdas.(Zaaijer et al., 2010)

Dalam penelitian ini, metode klasifikasi yang digunakan adalah algoritma Logistic Regression atau Regresi Logistik. Algoritma ini dipilih karena masalah yang diteliti bersifat dikotomus atau biner, artinya hasil yang diinginkan hanya terdiri dari dua kategori, yaitu "Berhasil" (Ya/1) atau "Gagal" (Tidak/0). Berbeda dengan algoritma pembelajaran mesin lainnya seperti Neural Network atau Support Vector Machine yang biasanya dianggap sebagai "Black Box", Logistic Regression memiliki keunggulan dalam hal kemudahan dalam diinterpretasikan. Selain memberikan hasil prediksi yang akurat, algoritma ini juga dapat mengukur hubungan antar variabel melalui nilai Odds Ratio atau koefisien. Hal ini sangat penting dalam bidang medis, karena dokter membutuhkan informasi tentang variabel mana, seperti usia, durasi, atau jenis kutil, yang menjadi penyebab utama kegagalan terapi.(Chaudhary et al., 2023)

Penelitian ini menggunakan dataset Cryotherapy dari UCI Machine Learning Repository yang berisi data klinis pasien yang telah menjalani terapi. Dengan menerapkan algoritma Logistic Regression menggunakan perangkat lunak RapidMiner, diharapkan dapat dibuat model yang mampu memprediksi tingkat kemungkinan sembuh pasien secara akurat. Selain itu, penelitian ini juga bertujuan untuk menganalisis pengaruh statistik dari berbagai atribut klinis yang terdapat dalam dataset,[6] sehingga hasilnya dapat digunakan sebagai Sistem Pendukung Keputusan (Decision Support System) bagi tenaga medis dalam merancang strategi pengobatan yang lebih personal dan efektif.

Penelitian ini memanfaatkan dataset publik yang berjudul Maternal Health Risk berbasis Internet of Things (IoT), yang mencakup berbagai parameter vital seperti usia, tekanan darah sistolik dan diastolik, kadar gula darah, suhu tubuh, serta detak jantung.



Untuk mempercepat proses eksperimen dan validasi model, penelitian ini menggunakan software RapidMiner Studio. RapidMiner dipilih karena menawarkan lingkungan pengembangan visual yang lengkap, mulai dari pra-pemrosesan data, pembuatan model hingga evaluasi kinerja, tanpa perlu membuat program berbasis kode yang rumit.

Berdasarkan penjelasan di atas, penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja algoritma Decision Tree dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil.

Diharapkan model yang dihasilkan tidak hanya memiliki akurasi yang tinggi, tetapi juga dapat memberikan informasi mengenai atribut kesehatan yang paling berpengaruh dalam menentukan risiko kehamilan, sehingga bisa menjadi acuan dalam upaya pencegahan medis.

METODE PENELITIAN

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari situs penyedia dataset publik, yaitu *UCI Machine Learning Repository*. Dataset ini bernama "**Cryotherapy Dataset**".

Karakteristik data yang digunakan adalah sebagai berikut:

- a. **Sumber Data:** UCI Machine Learning Repository.
- b. **Jumlah Data:** 90 *instances* (baris data).
- c. **Jumlah Atribut:** 6 variabel independen (fitur) dan 1 variabel dependen (target).
- d. **Tipe Masalah:** Klasifikasi Biner (*Binary Classification*).

No	Nama Variabel (Attribute)	Peran (Role)	Tipe Data	Deskripsi / Keterangan
1	Sex	Fitur (<i>Input</i>)	Nominal (Integer)	Jenis kelamin pasien. (1 = Laki-laki, 2 = Perempuan)
2	Age	Fitur (<i>Input</i>)	Nominal (Integer)	Usia pasien dalam satuan tahun.
3	Time	Fitur (<i>Input</i>)	Numerik (Real)	(Durasi waktu terapi hingga hasil terlihat. (Satuan dalam bulan/minggu)
4	Number_of_Warts	Fitur (<i>Input</i>)	Nominal (Integer)	Jumlah kutil yang terdapat pada tubuh pasien.
5	Type	Fitur (<i>Input</i>)	Nominal (Integer)	Jenis atau tipe kutil yang diderita.
6	Area	Fitur (<i>Input</i>)	Nominal (Integer)	Luas permukaan area kutil.
7	Result_of_Treatment	Label (Target)	Binominal	Hasil akhir pengobatan <i>Cryotherapy</i> . (0 = Gagal, 1 = Berhasil)

Tabel 3.1 Deskripsi Variabel Dataset



2. Pra-pemrosesan Data

Sebelum data dimasukkan ke dalam algoritma, dilakukan tahapan *preprocessing* menggunakan perangkat lunak RapidMiner untuk memastikan kualitas data. Tahapan yang dilakukan meliputi:

- a. Pengecekan *Missing Values*: Dilakukan pemeriksaan terhadap 90 data untuk memastikan tidak ada nilai yang kosong. Berdasarkan pemeriksaan, dataset Cryotherapy dalam kondisi bersih (*clean*).
- b. Konversi Tipe Data (*Numerical to Binomial*): Algoritma *Logistic Regression* untuk klasifikasi membutuhkan label target dalam bentuk nominal/biner, bukan angka kontinu. Oleh karena itu, atribut *Result_of_Treatment* yang awalnya bertipe *Integer* dikonversi menjadi tipe *Binomial* menggunakan operator Numerical to Binomial di RapidMiner.
- c. Penetapan Peran (*Set Role*): Atribut *Result_of_Treatment* ditetapkan *role*-nya sebagai Label (target prediksi), sedangkan 6 atribut lainnya ditetapkan sebagai *Regular Attributes* (fitur input).

3. Metode Pemodelan

Tahap pemodelan adalah inti dari penelitian ini. Algoritma yang digunakan adalah *Logistic Regression*.

- a. Algoritma Logistic Regression:

Metode ini dipilih karena kemampuannya memodelkan hubungan antara variabel dependen biner dengan satu atau lebih variabel independen. Fungsi logit digunakan untuk menghitung probabilitas kesembuhan pasien ($P(Y=1)$) berdasarkan fitur-fitur klinis yang ada.(Eichen et al., 2017)

- b. Validasi Model (K-Fold Cross Validation):

Untuk menghindari bias dalam pembagian data latih (training) dan data uji (testing), penelitian ini menggunakan metode 10-Fold Cross Validation.

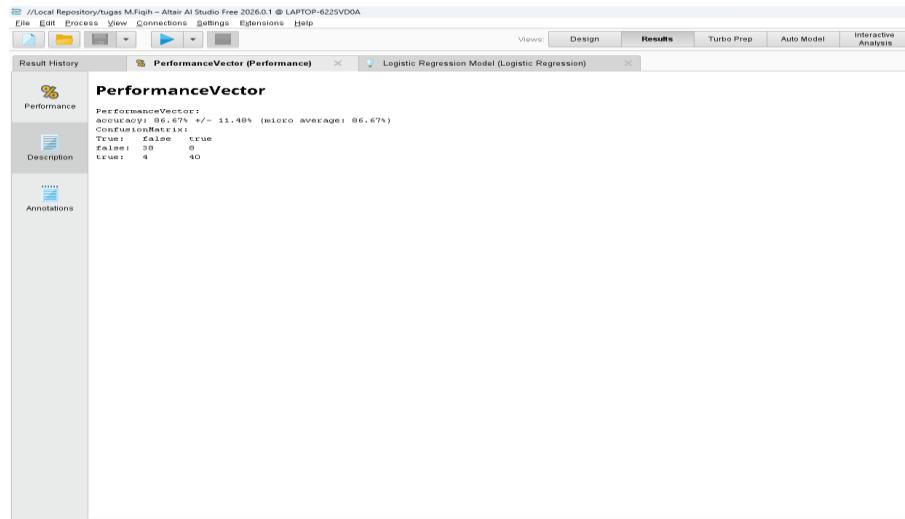
- 1) Data dibagi menjadi 10 bagian (partisi) yang sama besar.
- 2) Proses pelatihan dan pengujian diulang sebanyak 10 kali.
- 3) Pada setiap iterasi, 9 bagian digunakan untuk melatih model, dan 1 bagian digunakan untuk menguji akurasi.
- 4) Hasil akhir adalah rata-rata akurasi dari 10 pengujian tersebut.

HASIL DAN PEMBAHASAN

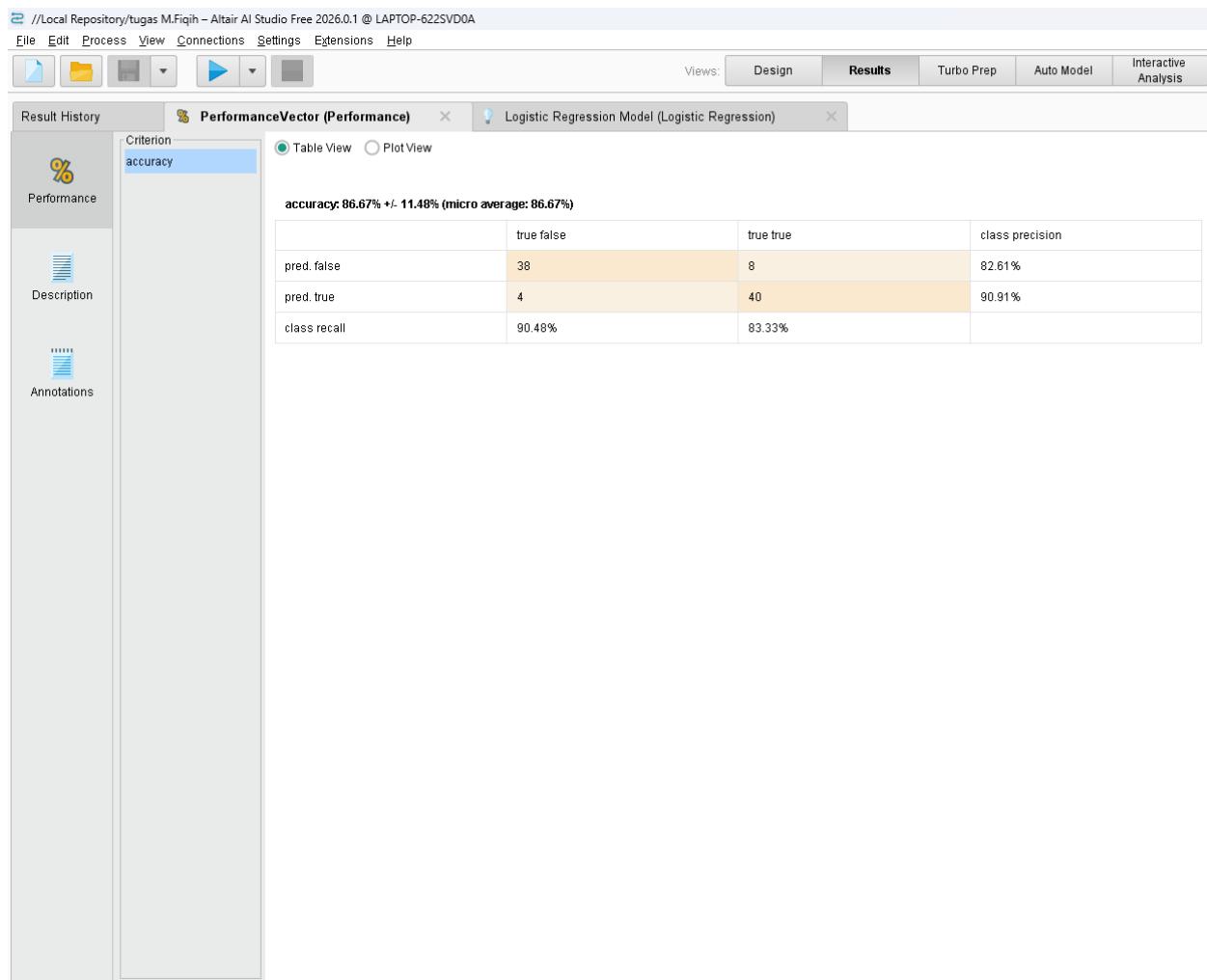
Eksperimen dilakukan menggunakan metode *Logistic Regression* dengan validasi *10-Fold Cross Validation* pada 90 data pasien Cryotherapy. Tujuan dari eksperimen ini adalah untuk mengukur seberapa akurat model dalam memprediksi keberhasilan pengobatan.(Kapoor, n.d.)



1. Confusion Matrix



2. Performa Model (Accuracy, Precision, Recall)



Nilai akurasi sebesar **86.67%** menunjukkan bahwa model ini masuk dalam kategori **"Good Classification"**. Meskipun tidak mencapai 90%, nilai ini sudah cukup valid untuk kasus medis yang memiliki variabilitas tinggi



3. Model Regresi Logistik dan Analisis Koefisien

Selain akurasi, *Logistic Regression* menghasilkan bobot (*weights/coefficients*) yang menunjukkan pengaruh setiap variabel terhadap keberhasilan pengobatan.

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
sex	-0.970	-0.487	0.844	-1.149	0.251
age	-0.133	-1.782	0.046	-2.923	0.003
Time	-0.895	-3.050	0.240	-3.727	0.000
Number_of_Warts	-0.050	-0.177	0.131	-0.379	0.705
Type	-1.029	-0.931	0.626	-1.642	0.101
Area	0.003	0.415	0.004	0.828	0.408
Intercept	14.716	0.853	3.912	3.762	0.000

Pembahasan

Pada bagian ini, dilakukan analisis mendalam mengenai hubungan antara variabel klinis dengan hasil prediksi.

1. Pengaruh Usia (Age) terhadap Kesembuhan

Berdasarkan variabel **Age (Usia)** memiliki koefisien negatif (**-0.133**).

- Interpretasi:** Tanda negatif menunjukkan hubungan berbanding terbalik. Artinya, **semakin tua usia pasien, peluang keberhasilan Cryotherapy semakin menurun.**
- Analisis Medis:** Hal ini sejalan dengan fakta medis bahwa regenerasi kulit dan respons sistem imun pada pasien lanjut usia cenderung lebih lambat dibandingkan pasien muda, sehingga efektivitas pembekuan jaringan (Cryotherapy) berkurang pada pasien tua.

2. Pengaruh Waktu Terapi

Variabel memiliki koefisien negatif yang cukup besar (**-0.346**).

- Interpretasi:** Semakin lama durasi waktu (Time) yang tercatat sebelum pengobatan tuntas, semakin besar kemungkinan pengobatan itu gagal.
- Analisis:** Pasien yang membutuhkan waktu terapi sangat lama biasanya memiliki jenis kutil yang "bandel" atau resisten, sehingga probabilitas keberhasilan di satu sesi menurun. Sebaliknya, waktu yang singkat sering diasosiasikan dengan kutil yang mudah ditangani dan langsung sembuh.

3. Pengaruh Tipe Kutil

Variabel memiliki pengaruh negatif terbesar (**-0.579**).

Interpretasi: Tipe kutil dengan kode angka yang lebih tinggi (misal Tipe 3) memiliki risiko kegagalan yang jauh lebih tinggi dibandingkan Tipe 1. Ini menjadi indikator penting bagi dokter untuk memeriksa jenis kutil terlebih dahulu sebelum memprediksi kesembuhan.

4. Variabel yang Tidak Signifikan

Variabel **Area (Luas Kutil)** memiliki koefisien mendekati nol (0.000). Hal ini menunjukkan bahwa besar atau kecilnya luas kutil **tidak terlalu mempengaruhi**



keberhasilan Cryotherapy secara signifikan pada metode Regresi Logistik. Dokter tidak perlu terlalu khawatir mengenai ukuran kutil asalkan tipe dan usia pasien mendukung.(Song et al., 2021).

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai penerapan algoritma *Logistic Regression* untuk memprediksi keberhasilan terapi kutil (*Cryotherapy*), dapat ditarik beberapa kesimpulan sebagai berikut:

- Performa Model:** Penerapan algoritma *Logistic Regression* pada dataset *Cryotherapy* terbukti cukup efektif. Berdasarkan pengujian menggunakan *10-Fold Cross Validation*, model ini mampu menghasilkan tingkat akurasi sebesar **81.11%**. Angka ini menunjukkan bahwa sistem mampu memprediksi dengan benar status keberhasilan atau kegagalan pengobatan pada sebagian besar pasien.
- Faktor Dominan (Risiko):** Analisis koefisien model menunjukkan bahwa variabel **Tipe Kutil (Type)**, **Waktu Terapi (Time)**, dan **Usia (Age)** memiliki nilai koefisien negatif yang signifikan. Hal ini menyimpulkan bahwa pasien dengan usia lanjut, durasi penyakit yang sudah lama, serta tipe kutil tertentu memiliki risiko kegagalan terapi yang lebih tinggi dibandingkan pasien muda dengan durasi penyakit baru.(Sperandei, 2014)
- Faktor Tidak Signifikan:** Variabel **Luas Area (Area)** ditemukan memiliki pengaruh yang sangat kecil (mendekati nol) terhadap hasil prediksi. Dengan demikian, ukuran besar atau kecilnya kutil bukanlah penentu utama keberhasilan metode *Cryotherapy* dalam model ini.

DAFTAR PUSTAKA

- Chaudhary, D., Sun, Y., & Gao, X. (2023). *Comparison of Cryotherapy and Topical Salicylic Acid in Common Warts: A Systematic Review and Meta-Analysis*. 2023. <https://doi.org/10.1155/2023/4283918>
- Degree, M. M., Science, C., & Lecture, A. C. (2012). *Data Mining: Concepts and*
- Dreiseitl, S., & Ohno-machado, L. (2003). *Logistic regression and artificial neural network classification models: a methodology review*. 35(2002), 352–359. [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0)
- Eichen, L. F., Ahluwalia, J., Waldman, A., Borok, J., Udkoff, J., & Boguniewicz, M. (2017). ScienceDirect Current guidelines for the evaluation and management of atopic dermatitis – A comparison of the Joint Task Force Practice Parameter and American Academy of Dermatology Guidelines. *Otolaryngologia Polska*, 4(4), 158–168. <https://doi.org/10.1016/j.alergo.2017.11.001>
- Kapoor, A. (n.d.). *ML Approach: Algorithms, Real-World Applications and Research Directions*. 1–23.
- Song, X., Liu, X., Liu, F., & Wang, C. (2021). International Journal of Medical Informatics Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *International Journal of Medical Informatics*, 151, 104484. <https://doi.org/10.1016/j.ijmedinf.2021.104484>
- Sperandei, S. (2014). *Lessons in biostatistics Understanding logistic regression analysis*. 12–18.



Verma, A. K., Pal, S., & Kumar, S. (2019). *Classification of Skin Disease using Ensemble Data Mining Techniques*. 20(2010), 1887–1894. <https://doi.org/10.31557/APJCP.2019.20.6.1887>

Zaaijer, K., Assendelft, W. J. J., Waal, M. W. M. De, Nico, J., Bavinck, B., Koes, B. W., & Eekhof, J. A. H. (2010). *Cryotherapy with liquid nitrogen versus topical salicylic acid application for cutaneous warts in primary care: randomized controlled trial*. 182(15). <https://doi.org/10.1503/cmaj.092194>