

Prediksi Diagnosa Berdasarkan Data Rekam Medis Pasien Menggunakan Support Vector Regression

Muhammad Nasrul Tsalatsa Putra^{*1}, Agus Eko Minarno², Setio Basuki³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang
Muhammad.nasrul228@gmail.com^{*1}, agoes.minarno@gmail.com²,
setiobasuki.umm@gmail.com³

Abstrak

Rekam medis merupakan suatu berkas dari hasil pemeriksaan kesehatan, pengobatan yang diberikan, tindakan, dan pelayanan lain yang telah diberikan kepada pasien. Penelitian ini dilandasi oleh beberapa permasalahan, diantaranya (1) kurangnya pengawasan, informasi, dan tidak meratanya pemberian layanan kesehatan, (2) terhambatnya perencanaan puskesmas dalam menanggulangi kasus yang sudah ada atau yang sering terjadi karena tingginya jumlah dan keberagaman kasus/diagnosa yang ditemukan di masyarakat. Dari permasalahan tersebut dapat diterapkan sistem prediksi diagnosa dengan menerapkan metode Support Vector Regression (SVR). Model SVR yang diterapkan yaitu kernel Linear, kernel Polynomial, serta kernel Radial Basis Function. Pengujian dilakukan dengan membagi dataset ke dalam data uji dan data latih, kemudian dilakukan proses pengujian hingga 9-fold untuk masing-masing model dengan susunan data yang berbeda. Hasil pengujian menunjukkan fungsi kernel RBF memiliki kinerja terbaik dibanding dengan fungsi lainnya dimana nilai NRMSE tertinggi 0.0797 dan nilai akurasi terendah sebesar 0.4826. Hasil prediksi tersebut dapat memberikan sebuah gambaran dan trend mengenai diagnosa yang akan datang berdasarkan data rekam medis pasien.

Kata Kunci: Prediksi Diagnosa, Support Vector Regression, Kernel, Rekam Medis

Abstract

Medical record is a file of health examination result, medication given, along with treatment and other service given to the patient. This research is based on a few problems, which are: (1) lack of supervising, information, and uneven distribution of health service, (2) delay of health center planning on treating already existing or often occurring case because the number and the variety of case/final diagnose found in society is quite high. From these problems, can be applied a diagnose prediction system is using Support Vector Regression (SVR) method. SVR models used are kernel Linear, kernel Polynomial, and kernel Radial Basis Function. The test is done by dividing dataset into test data and training data, therefore will be conducted process of testing can be done up to 9-fold for each models with different data alignment. Test result showed kernel RBF function has the best performance among other functions which mean value of NRMSE is 0.0797 and mean value of accuracy is 0.4826. That prediction result can give an illustration and a trend about upcoming diagnose based on patient medical record data.

Keywords: Diagnose Prediction, Support Vector Regression, Kernel, Medical Records

1. Pendahuluan

Upaya meningkatkan kesehatan dalam masyarakat memiliki tujuan meningkatkan nilai kesehatan dan pemerataan pelayanan kesehatan [1]. Tingkat penilaian kesehatan memerlukan upaya untuk meningkatkan nilai kesehatan dalam masyarakat [2]. Beberapa permasalahan dalam Upaya meningkatkan nilai kesehatan yakni puskesmas selalu disibukkan dengan administrasi berupa laporan keuangan, laporan kesehatan perbulan, serta permasalahan lainnya. Permasalahan tersebut berimbas pada efisiensi bantuan kesehatan bila lembaga tersebut juga harus mencukup administrasi setiap bulannya. Dengan permasalahan yang ada tersebut, dapat diterapkan penggunaan teknologi dimana terdapat beberapa terobosan baru dalam bidang kesehatan. Terobosan teknologi kesehatan yaitu sistem informasi. Sistem tersebut tercatat data berkas pemeriksaan pasien [3].

Berkas pemeriksaan tersebut akan bermanfaat bila digunakan untuk pemetaan wilayah dan pengelolaan data. Pada permasalahan kurangnya pengawasan, informasi, dan tidak

meratanya bantuan kesehatan, maka diberikan solusi yaitu sebuah sistem pemetaan wilayah (sesuai dengan daerah puskesmas). Sistem tersebut memberikan sebuah informasi terkait pasien yang memiliki diagnosa tertentu sesuai dengan desa asal pasien. Dalam sistem tersebut, dapat diketahui persebaran pasien dengan diagnosa tertentu seperti diagnosa influenza, diabetes mellitus, dan lainnya. Pemetaan wilayah akan menerapkan heatmap layer. Pada penelitian saudara Endah Dharmaputeri mengenai pemetaan wilayah kesehatan menggunakan Quantum GIS menyatakan penelitian tersebut melakukan visualisasi wilayah dengan arsip pemeriksaan pasien untuk acuannya [4].

Pada permasalahan lainnya yaitu terhambatnya perencanaan puskesmas dalam menanggulangi kasus yang sudah ada atau yang sering terjadi karena tingginya jumlah dan keberagaman kasus/diagnosa. Dari permasalahan tersebut diberikan solusi yaitu sebuah sistem yang akan memprediksi diagnosa dengan mengacu pada data rekam medis pasien. Prediksi merupakan suatu kegiatan peramalan untuk memperoleh nilai mendatang dengan mempertimbangkan kejadian / data sebelumnya. Berbagai metode dapat mengolah data untuk memperoleh hasil prediksi, salah satu yakni metode Support Vector Regression (SVR) dimana SVR dapat mengatasi overfitting dan kasus non-linear sehingga akan mendapatkan nilai yang hampir sama dengan data ujinya [3]. Pada penelitian saudara Retno Larasati mengenai Support Vector Regression menyatakan bahwa model kernel Radian Basis Function memiliki kinerja terbaik dengan nilai akurasi 0.7 dan RMSE 2.8 dengan memanfaatkan data SOI (Southern Oscillation Index) dan data observasi dari stasiun cuaca di Indramayu [5].

Pada penelitian saudara Dwi Otik Kurniawati dkk dengan topik kesehatan. Pada penelitian tersebut menggunakan data rekam medis dan pemeriksaan laboratorium untuk menentukan suatu diagnosa dengan sistem neuro fuzzy menyatakan dengan adanya learning algorithm pada sistem kesehatan diharapkan sistem dapat melakukan pembelajaran sehingga keluaran yang didapatkan mendekati kebenaran [6].

Merujuk pada permasalahan dan penelitian sebelumnya, maka dilakukan implementasi Support Vector Regression untuk mendapatkan hasil prediksi diagnosa. Selain itu visualisasi mengenai persebaran pasien dilakukan dengan sistem pemetaan wilayah berdasarkan berkas pemeriksaan pasien sesuai dengan kondisi wilayah Turen.

2. Metode Penelitian

2.1. Data Penelitian

Penelitian ini menggunakan data dari Puskesmas Turen, Kabupaten Malang. Data yang digunakan telah mendapatkan izin dari Badan Kesatuan Bangsa dan Politik Kabupaten Malang. Data yang didapatkan berupa data master pasien, 14700 data diagnosa, data master desa, dan 2700 data rekam medis poli anak dan poli ibu pada periode bulan Agustus 2016 hingga bulan April 2017.

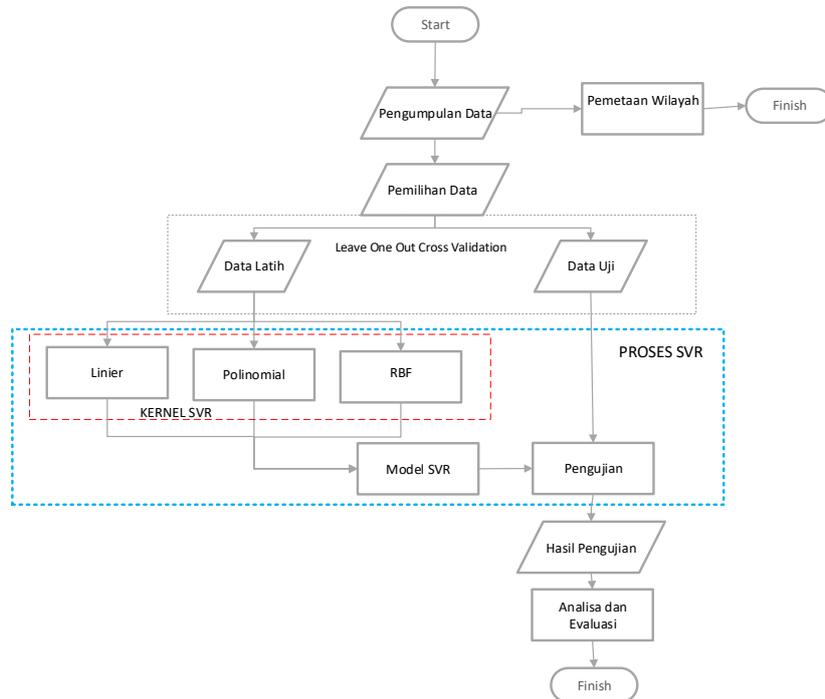
Data pada poli anak dan poli ibu digunakan untuk membuat sebuah grafik dan visualisasi wilayah. Namun pada fitur prediksi hanya digunakan data pada poli anak untuk memprediksi diagnosa dengan menggunakan metode SVR. Seluruh data yang didapatkan bisa langsung digunakan tanpa perlu dilakukan preprocessing.

2.2. Pemetaan Wilayah dan Prediksi Diagnosa

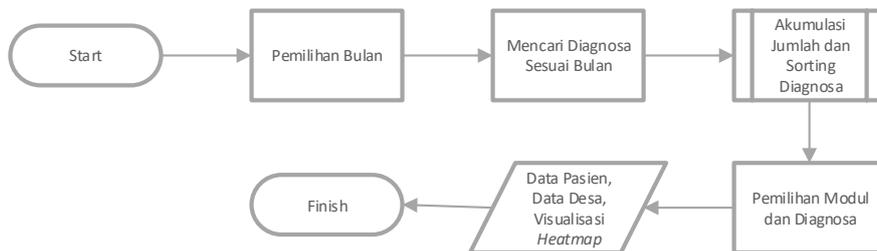
Pada penelitian ini dibangun sistem pemetaan wilayah dan prediksi diagnosa berdasarkan permasalahan yang sudah dijabarkan. Pada pemetaan wilayah menggunakan hasil pemeriksaan poli anak dan poli ibu namun akan diterapkan grafik 15 diagnosa terbanyak setiap bulannya, sehingga *user* memilih diagnosa dari grafik tersebut untuk melihat persebaran pasien pada daerah tersebut. Sedangkan untuk prediksi diagnosa menggunakan hasil pemeriksaan poli anak. Prediksi prediksi tersebut dapat memberikan sebuah gambaran dan trend mengenai diagnosa yang akan datang berdasarkan data rekam medis pasien. Gambar 1 merupakan diagram alir yang dilakukan pada penelitian.

Data rekam medis pasien didapatkan dari Puskesmas Turen berupa data rekam medis dari poli anak dan poli ibu. Data dikoleksikan berdasarkan kategorinya. Fitur pemetaan wilayah akan menggunakan data yang telah terkumpul. *User* akan memilih bulan dengan rentang Agustus 2016 hingga April 2017. Setelah *user* memilih bulan, maka sistem akan menampilkan grafik diagnosa dimana grafik tersebut berdasarkan 15 penyakit terbanyak perbulannya yang sudah dijumlahkan. Setelah sistem berhasil menampilkan grafik, maka *user* dapat memilih salah satu kategori modul layanan dan salah satu diagnosa untuk mengetahui persebaran pasien dengan diagnosa yang

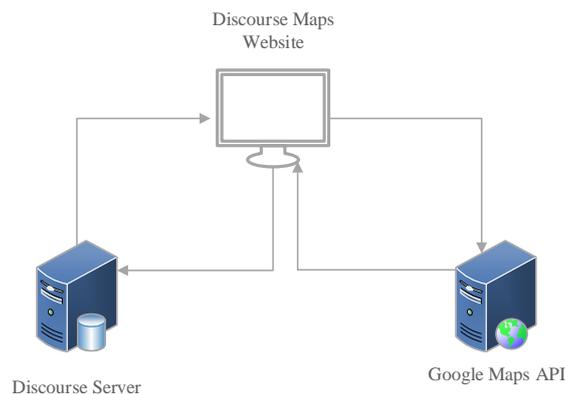
di pilih. Sistem akan menampilkan data pasien, data desa, dan *heatmap* berdasarkan diagnosa yang di pilih oleh user. Dapat di lihat pada Gambar 2 merupakan alur visualisasi wilayah.



Gambar 1. Alur Sistem Pemetaan Wilayah dan Prediksi Diagnosa



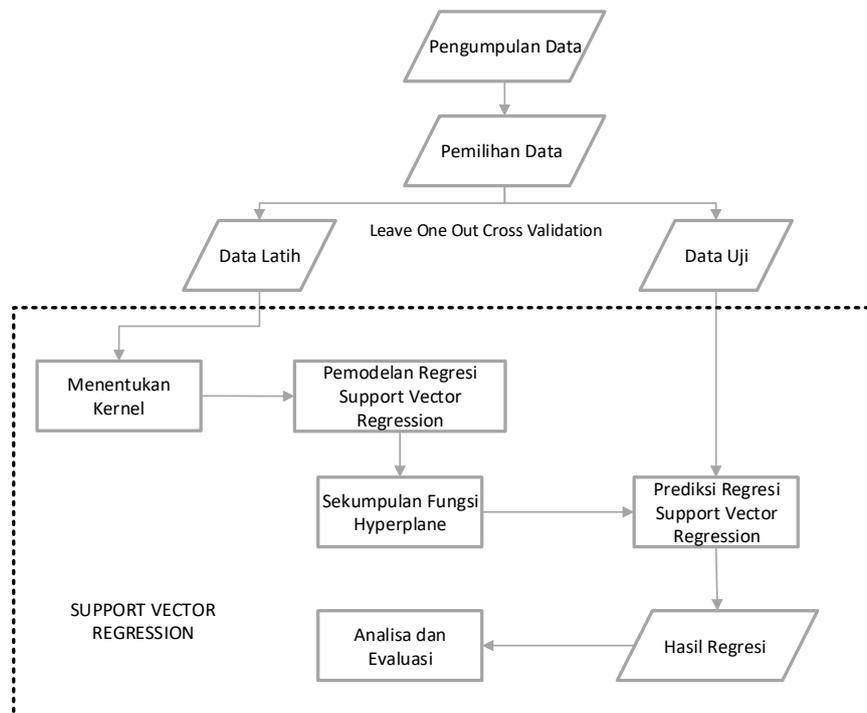
Gambar 2. Alur Sistem Pemetaan Wilayah



Gambar 3. Arsitektur Maps

Lokasi untuk masing-masing desa akan diambil dengan menggunakan latitude dan longitude, dimana tersimpan pada discourse server. Ketika memvisualisasikan desa tersebut maka, latitude dan longitude tersebut akan dikirimkan ke server Google Maps API sehingga akan didapatkan sebuah gambaran berupa maps sesuai dengan *latitude* dan *longitude* yang akan dimunculkan pada website. Arsitektur maps dapat di lihat pada Gambar 3.

Dalam hal prediksi data yang dikumpulkan dilakukan pemilihan data. Pemilihan data tersebut akan di bagi menjadi dua jenis, yaitu data latih dan data uji. Untuk pembagian data menggunakan *Leave One Out Cross Validation* dimana data yang telah dikumpulkan tersebut dibagi secara *random*. Pembagian data akan terbagi ke dalam data latih sebesar 2400 data, sedangkan data uji sebesar 300 data. Setelah pembagian data dilakukan, data latih digunakan untuk proses pembangunan model dari metode SVR. Fungsi model yang digunakan yaitu kernel Linear, Polynomial, dan Radial Basis Function. Model SVR yang telah dibangun siap untuk dilakukan pengujian.



Gambar 4. Alur Kerja Support Vector Regression

Pada tahap pengujian, data uji akan digunakan sebagai masukkan untuk mendapatkan sebuah hasil prediksi dari model SVR yang telah dibangun. Setelah dilakukan pengujian maka didapatkan hasil prediksi. Pengujian akan dilakukan sebanyak *9-fold* untuk masing-masing model SVR. Pengukuran kinerja model SVR dilakukan dengan mencari nilai akurasi dan NRMSE. Pengukuran kinerja model SVR tersebut digunakan untuk mengevaluasi dan menganalisa terhadap model SVR yang telah dibangun sehingga didapatkan nilai akurasi tertinggi dengan nilai yang mendekati angka 1 dan nilai NRMSE terendah dengan nilai yang mendekati angka 0. Alur kerja Support Vector Regression dapat di lihat pada Gambar 4.

2.3. Analisa dan Evaluasi Model Support Vector Regression

Dalam kasus regresi diperlukannya sebuah analisa dan evaluasi terkait pengujian yang telah dilakukan. Terdapat nilai akurasi sering dipakai untuk menilai performansi keberhasilan model SVR. Selain itu nilai *Normalized Root Mean Squared Error* (NRMSE) digunakan untuk mengevaluasi tingkat error terhadap pengujian. Dengan percobaan pengujian yang telah dilakukan, maka akan dicari model terbaik yang menghasilkan hasil prediksi yang mendekati nilai aslinya. Akan diperoleh nilai akurasi tertinggi dengan nilai yang mendekati angka 1 dan nilai NRMSE terendah dengan nilai yang mendekati angka 0, seperti pada Persamaan 1 dan Persamaan 2.

Normalized Root Mean Squared Error (NRMSE)

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$y_{max} - y_{min}$$

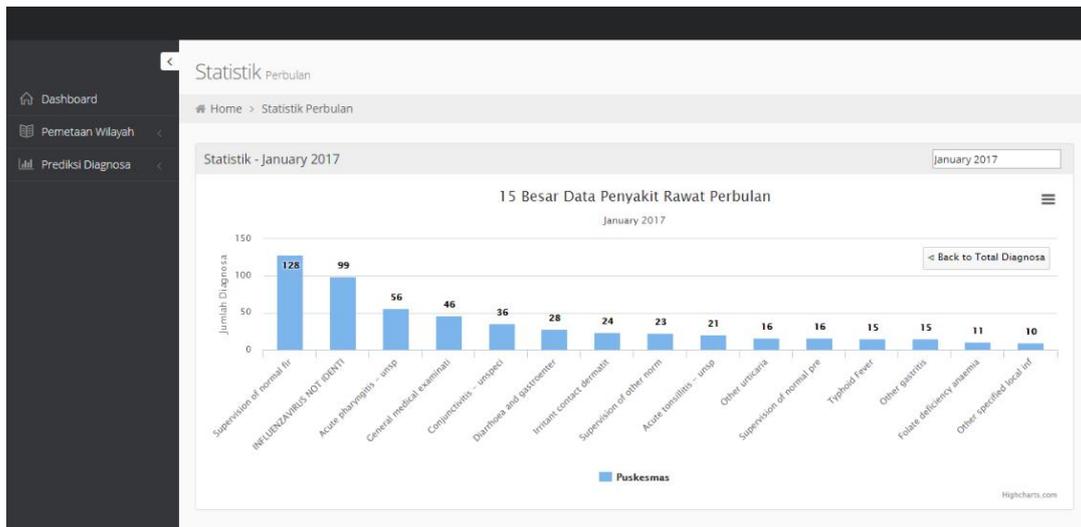
Akurasi

$$\frac{fx}{fn} \tag{2}$$

3. Hasil Penelitian dan Pembahasan

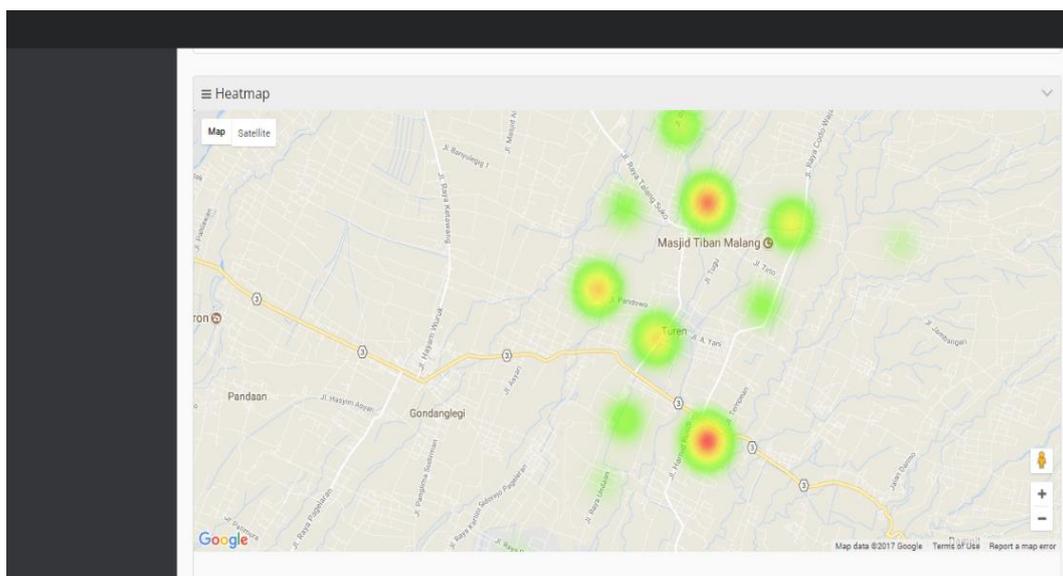
3.1. Penerapan Pemetaan Wilayah

Data yang telah didapatkan dan dikumpul akan diterapkan fitur pemetaan wilayah dengan menggunakan *heatmap layer*. Data yang didapatkan diterapkan sebuah grafik mengenai 15 diagnosa terbanyak perbulannya seperti Gambar 5.



Gambar 5. Grafik 15 Besar Penyakit Terbanyak

User memilih diagnosa pada grafik tersebut untuk dapat melihat persebaran pasien pada daerah tersebut. Persebaran tersebut hanya sesuai data desa pada data master desa. Pada data master sudah terdapat *latitude* dan *longitude*. Akan ditampilkan visualisasi berupa heatmap dari diagnosa yang di pilih tersebut [8]. Heatmap tersebut mengambil lokasi desa (latitude dan longitude) berdasarkan asal desa pasien yang memiliki diagnosa yang sama dengan yang di pilih. Semakin pekat warna pada heatmap, maka pada lokasi tersebut memiliki jumlah diagnosa yang tinggi. Sebaliknya, bila warnanya tidak terang / tidak pekat maka jumlah diagnosa semakin kecil [9].



Gambar 6. Heatmap Layer

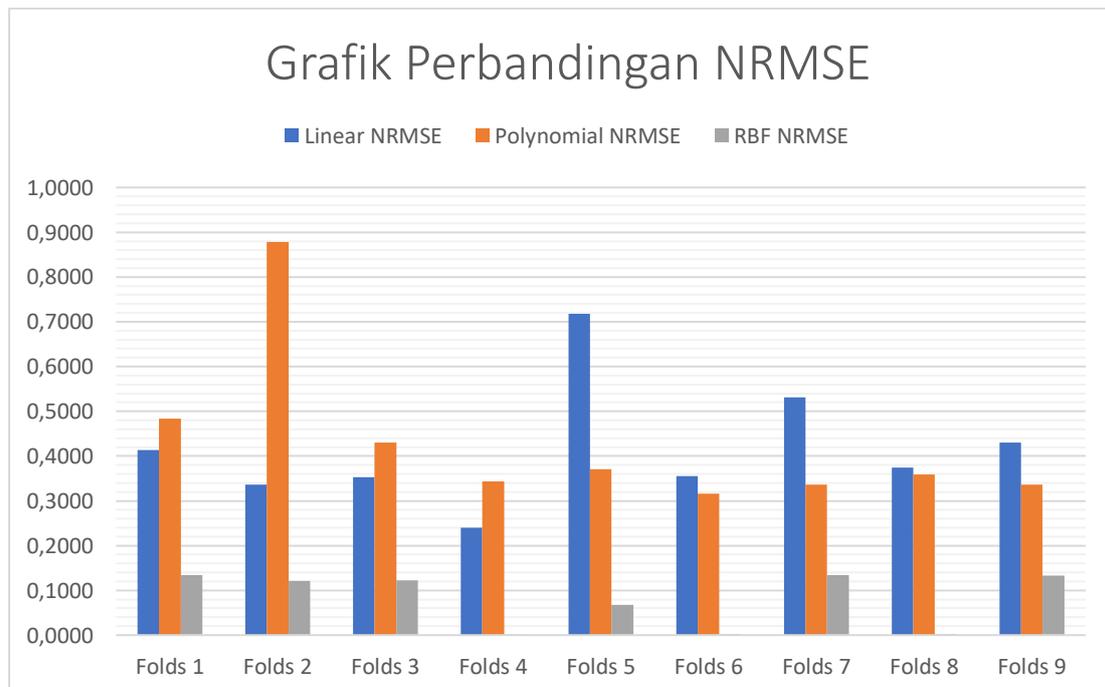
3.2. Pengujian Kinerja Model SVR

Data yang didapatkan akan dilatih dan diuji dengan fungsi kernel linear, kernel polynomial, dan kernel RBF. Pengujian dilakukan pada model SVR yang dihasilkan dengan metode *LOO 9-fold Cross Validation*. Kinerja dari luaran dapat dilihat berdasarkan nilai *error* dan akurasinya. Luaran dikatakan memiliki kinerja terbaik apabila nilai *error* yang dihasilkan kecil atau nilai *error* mendekati angka 0 dan nilai akurasinya besar atau mendekati angka 1 [10]. Hasil dari pelatihan dengan menggunakan *9-fold* untuk masing-masing kernel dapat dilihat pada Tabel 1. Pada Tabel 1 tersebut memiliki nilai NRMSE dan akurasi yang bervariasi. Namun pada percobaan *9-folds* pada kernel RBF memiliki nilai NRMSE dan nilai akurasi yang paling sesuai dimana nilai NRMSE mendekati angka 0 dan nilai akurasi mendekati angka 1.

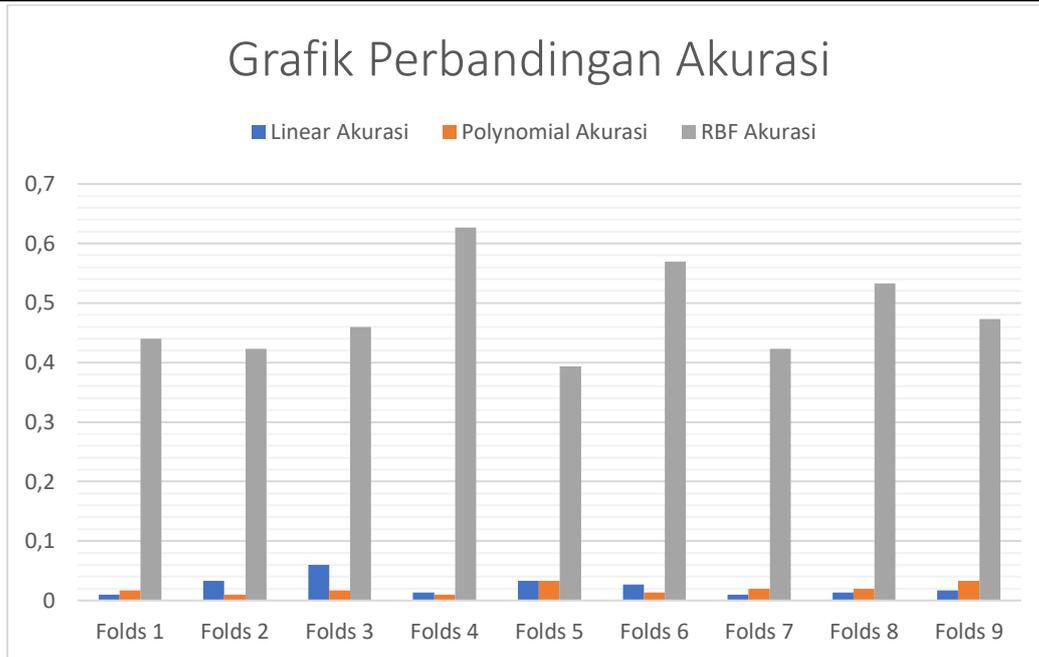
Tabel 1. Tabel Perbandingan NRMSE dan Akurasi

	Linear		Polynomial		RBF	
	NRMSE	Akurasi	NRMSE	Akurasi	NRMSE	Akurasi
Folds 1	0.4137	0.0100	0.4832	0.0167	0.1344	0.4400
Folds 2	0.3360	0.0333	0.8780	0.0100	0.1212	0.4233
Folds 3	0.3524	0.0600	0.4301	0.0167	0.1218	0.4600
Folds 4	0.2402	0.0133	0.3439	0.0100	0.0017	0.6267
Folds 5	0.7172	0.0333	0.3710	0.0333	0.0674	0.3933
Folds 6	0.3551	0.0267	0.3161	0.0133	0.0018	0.5700
Folds 7	0.5312	0.0100	0.3359	0.0200	0.1342	0.4233
Folds 8	0.3742	0.0133	0.3584	0.0200	0.0021	0.5333
Folds 9	0.4301	0.0167	0.3360	0.0333	0.1328	0.4733

Dari data tersebut bisa digambarkan dengan sebuah grafik yang menunjukkan hasil penelitian untuk seluruh percobaan pada masing-masing model SVR. Bila diperhatikan pada Gambar 7 kernel RBF paling rendah dibandingkan kernel lainnya dan Gambar 8 kernel RBF paling tinggi dibandingkan kernel lainnya. Sedangkan kernel linear dan polynomial memiliki pergerakan grafik secara fluktuatif.



Gambar 7. Grafik Perbandingan NRMSE



Gambar 8. Grafik Perbandingan Nilai Akurasi

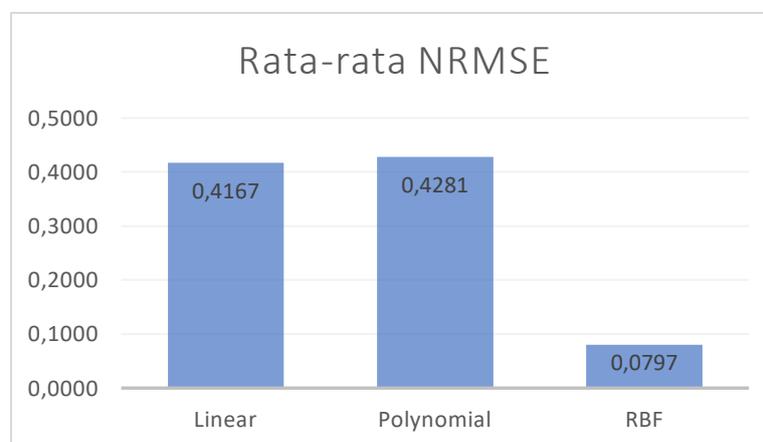
Hasil dari pelatihan dapat di lihat pada Tabel 2. Nilai rata-rata akurasi tertinggi didapat dengan menggunakan fungsi kernel RBF dengan nilai 0.4826.

Tabel 2. Hasil Uji Model dengan SVR

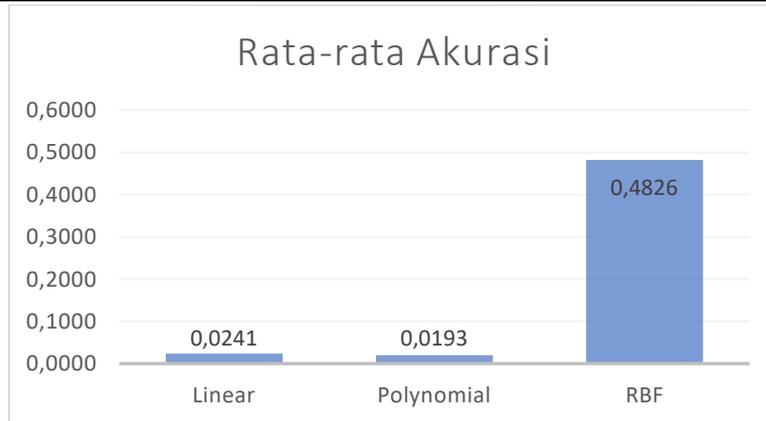
	Linear	Polynomial	RBF
Rata-rata NRMSE	0.4167	0.4281	0.0797
Rata-rata Akurasi	0.0241	0.0193	0.4826

Rata-rata nilai normalisasi akar error rata-rata kuadrat (NRMSE) pada data rekam medis tersebut menunjukkan hasil sesuai dengan akurasinya. Karena akurasi dan NRMSE memiliki hasil yang terbalik. Model dengan nilai akurasi tertinggi memiliki NRMSE terendah yaitu model dengan kernel RBF, yakni sebesar 0.0797. Nilai NRMSE tertinggi didapat dari model dengan kernel polynomial. Hal ini dikarenakan hasil prediksi pada kernel polynomial sangat jauh berbeda dengan data ujinya.

Hasil prediksi diagnosa memperlihatkan variasi ukuran *error* dan akurasi. Nilai *error* dengan menggunakan fungsi kernel RBF menghasilkan NRMSE terendah sebesar 0.0797 dengan nilai akurasinya sebesar 0.4826 untuk simulasi dengan menggunakan LOO 9-fold cross validation.

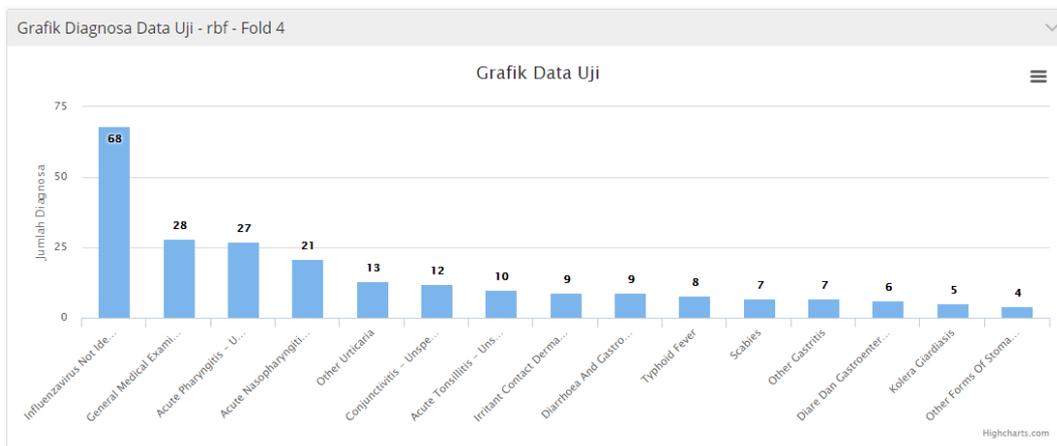


Gambar 9. Perbandingan Rata-rata Nilai NRMSE

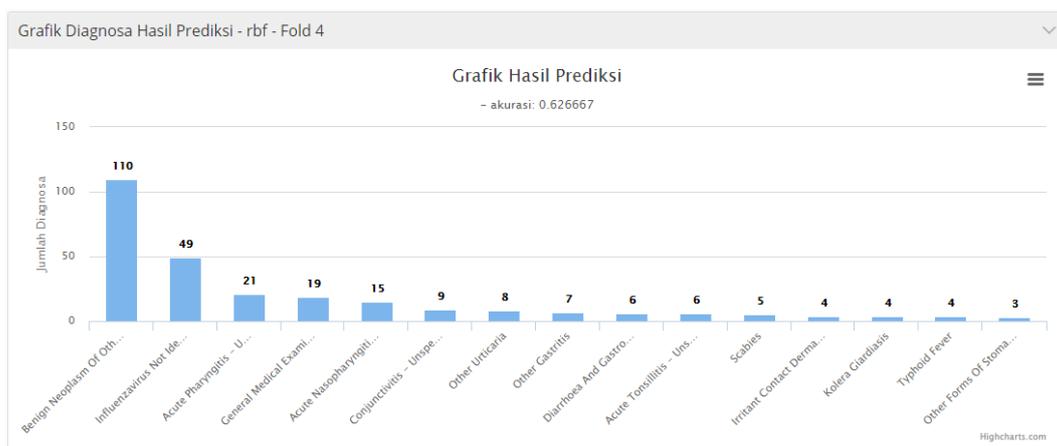


Gambar 10. Perbandingan Rata-rata Nilai Akurasi

Gambar 9 dan Gambar 10 merupakan grafik data uji dan hasil prediksi diagnosa dari model radial basis function pada *fold* ke 4. Gambar tersebut menunjukkan bahwa dari grafik tersebut dapat memberikan sebuah gambaran dan trend mengenai diagnosa yang akan datang berdasarkan data rekam medis pasien.



Gambar 11. Grafik 15 Penyakit dari Data Uji Kernel RBF Fold 4



Gambar 12. Grafik 15 Penyakit dari Hasil Prediksi Kernel RBF Fold 4

4. Kesimpulan

Berlandaskan pengujian memperoleh kesimpulan dimana fungsi model kernel RBF memiliki kinerja terbaik dibanding dengan model fungsi kernel lainnya. Pada pengujian model SVR tersebut mendapatkan nilai NRMSE tertinggi dan nilai akurasi terendah pada fungsi kernel

RBF. Hasil prediksi yang dihasilkan dapat menyampaikan sebuah gambaran dan trend mengenai diagnosa yang akan datang berdasarkan berkas pemeriksaan pasien. Adapun saran yang penulis berikan untuk penelitian berikutnya yaitu pada model SVR sebaiknya menentukan nilai parameter yang lebih sesuai dengan kernel-kernelnya sehingga model SVR yang dibangun memperoleh hasil prediksi yang mendekati data aslinya.

Referensi

- [1] M. Tiyas and A. Rohmani, "Hubungan Kepuasan Pasien Dengan Minat Pasien Dalam Pemanfaatan Ulang Pelayanan Kesehatan Pada Praktek Dokter Keluarga," *Fak. Kedokt. Univ. Muhamadiyah Semarang*, vol. 1, no. 1, p. 155, 2012.
- [2] D. Rachmat, "Implementasi Metode Sorensen Coefficient Dalam Menentukan Daerah Berpotensi Rawan Penyakit Demam Berdarah (Studi Kasus : Kota Pontianak)."
- [3] G. Susanto and Sukadi, "Sistem Informasi Rekam Medis Pada Rumah Sakit Umum Daerah (Rsud) Pacitan Berbasis Web Base," *Speed – Sentra Penelit. Eng. dan Edukasi*, vol. 9, no. 3, pp. 40–46, 2012.
- [4] E. Dharmaputeri, "Aplikasi Sistem Informasi Geografis Pelayanan Kesehatan Kota Depok Berbasis Web Menggunakan Quantum Gis," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.
- [5] H. Gunawan, "Seleksi Hyperspectral Band Menggunakan Recursive Feature Elimination Untuk Prediksi Produksi Padi Dengan Support Vector Regression," 2012.
- [6] Retno Larasati, "Prediksi Awal Musim Hujan Menggunakan Data Southern Oscillation Index Dengan Metode Support Vector Regression," 2012.
- [7] D. O. Kurniawati *et al.*, "Diagnosis Penyakit Pasien Menggunakan Sistem Neuro Fuzzy," vol. 2014, no. Sentika, 2014.
- [8] Siswanto, "Sistem Informasi Geografis Objek Wisata Menggunakan Google Maps Api Studi Kasus Kabupaten Mojokerto," pp. 1–5.
- [9] F. H. Suryanto, "Perancangan Aplikasi Pemetaan Persebaran UKM yang Memproduksi Makanan dan Minuman di Salatiga Menggunakan Heatmap," *Univ. Kristen Satya Wacana*, pp. 2–17, 2014.
- [10] H. Yasin, A. Prahutama, and T. Wahyu, "Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search," *Media Stat.*, vol. 7, pp. 29–35, 2014.

