

PERAMALAN TIME SERIES MENGGUNAKAN GAUSSIAN KERNEL PCA DAN AUTOREGRESSIVE

Kasiful Aprianto¹

Badan Pusat Statistik Provinsi Sulawesi Barat
e-mail: ¹apriantokasiful@gmail.com

Abstrak

Peramalan time series menjadi bagian penting dalam pengambilan keputusan karena mampu memberikan gambaran atau kejadian mendatang berupa data prediksi. Paper ini menawarkan gaussian kernel PCA dan Autoregressive (KPCA-AR) sebagai metode peramalan untuk melakukan prediksi data. Metode KPCA-AR bekerja dengan melakukan pemetaan data ke dimensi yang lebih tinggi menggunakan kernel dengan distribusi gaussian. Setelah itu dilakukan transformasi data dengan PCA agar dimensi yang dihasilkan dapat direduksi dengan varian maksimum sehingga tidak mengurangi karakteristik data secara signifikan. Data inilah yang kemudian digunakan untuk melakukan peramalan menggunakan autoregressive. Paper ini juga membandingkan beberapa metode peramalan lainnya seperti ARIMA, ANN, SVM, dan Eksponensial Smoothin. Hasil menunjukkan bahwa KPCA-AR secara umum mampu memberikan prediksi yang baik dan bisa digunakan sebagai alternative dari metode perhitungan yang ada dilihat dari kelebihan ataupun kekurangannya.

Kata kunci: *Gaussian, Kernel, PCA, Autoregressive, Forecasting*

Abstract

Time series forecasting becomes an important part in decision making because it is able to provide a picture or future event in the form of predictive data. This paper offers the Gaussian kernel PCA and Autoregressive (KPCA-AR) as a forecasting method to predict data. The KPCA-AR method works by mapping data to a higher dimension using a kernel with a gaussian distribution. After that the transformation of data with PCA is done so that the resulting dimensions can be reduced with maximum variants so as not to significantly reduce the characteristics of the data. This data is then used to forecast using autoregressive. This paper also compares several other forecasting methods such as ARIMA, ANN, SVM, and Exponential Smoothing as a comparison. The results show that KPCA-AR in general is able to give good predictions and can be used as an alternative to existing calculation methods seen from the advantages or disadvantages.

Keywords: *Gaussian, Kernel, PCA, Autoregressive, Forecasting*

PENDAHULUAN

Peramalan yaitu kegiatan memprediksi sesuatu yang akan terjadi kedepan dengan berbagai pertimbangan dan kemungkinan berdasarkan fakta yang terjadi hari ini dan sebelumnya. Dengan memprediksi kejadian mendatang menggunakan data historis yang tersedia, pemilik kepentingan mampu mendapatkan gambaran umum kedepan dan menetapkan kebijakan agar tujuan tepat sasaran. Peramalan time series telah banyak digunakan untuk beberapa kasus prediksi seperti pertanian, iklim, ekonometrik, finansial, dan komunikasi.

Beberapa teknik peramalan yang dilakukan sebelumnya seperti SVM, Neural Network, exponential smoothing, ARIMA, dan beberapa teknik lainnya telah berhasil memberikan prediksi yang baik sesuai pada percobaan yang pernah dilakukan [1][2]. Selanjutnya dilakukan sejumlah percobaan seperti penggabungan beberapa metode untuk meningkatkan akurasi peramalan. Diantaranya seperti ARIMA-ANN, dimana dilakukan penggabungan antara metode ARIMA dan Neural network [3], lalu penggabungan metode algoritma genetika dan SVM [4], dimana kedua hasil memberikan peningkatan akurasi. Pada penelitian ini penulis akan melakukan kombinasi antara Gaussian kernel, PCA, dan autoregressive, untuk selanjutnya akan disingkat dengan KPCA-AR.

KPCA-AR bertujuan untuk menghilangkan asumsi normalitas suatu data sehingga bisa diprediksi menggunakan autoregressive, dimana syarat agar bisa dilakukan peramalan dengan metode autoregressive yaitu data terdistribusi secara normal. Hal tersebut dapat diatasi dengan melakukan transformasi data dari linear menjadi nonlinear. Untuk beberapa kasus, jika suatu data tidak terpenuhi asumsi normalitas, maka harus di transformasi, berupa polynomial, log, atau berbagai bentuk lainnya. Salah satunya yaitu dengan menggunakan kernel. Kernel sendiri merupakan teknik transformasi dengan melakukan proyeksi data ke dimensi yang lebih tinggi. Teknik ini juga

digunakan pada SVM sehingga data mampu dipisah secara linear dengan dimensi yang berbeda. Kenyataannya, dimensi yang dihasilkan kernel cenderung besar dan membuat proses komputasi semakin besar. Untuk itu dilakukan pengurangan dimensi tanpa mengurangi makna atau karakteristik dari data tersebut secara signifikan. Salah satu metodenya adalah PCA. PCA sendiri merupakan teknik menyederhanakan data dengan transformasi yang mampu membentuk koordinat baru dengan varian maksimum.

Gaussian Kernel

Kernel adalah suatu cara melakukan pemetaan data dari data ruang inisial ke ruang fitur. Salah satu contoh fungsi kernel adalah kernel gaussian.

$$k(\bar{x}, \bar{x}') = \exp \left(\frac{-(\bar{x} - \bar{x}')^2}{2\sigma^2} \right)$$

dengan $\bar{x}$ dan $\bar{x}'$ adalah suatu data vektor yang menjelaskan sebuah titik dengan dimensi n . Terlihat bahwa $(\bar{x} - \bar{x}')$ merefleksikan jarak antar titik, dan σ merupakan suatu variabel bebas.

Misalkan $A = \{\bar{x}^1 \dots \bar{x}^m\}$ dimana dataset A terdiri dari titik ke-1 sampai titik ke- m , masing-masing berdimensi n vector, maka fungsi K dipetakan seperti berikut.

$$k(\bar{x}^i, \bar{x}^j): \mathbb{R}^n \cdot \mathbb{R}^n \rightarrow \mathbb{R}$$

Fungsi di atas merupakan kernel jika dan hanya jika matriks yang dihasilkan adalah matriks Gram yang dapat dihasilkan dari inner product. Sebagai bukti, misal fungsi kernel untuk titik i dan j , dapat dijelaskan dalam rumus berikut:

$$\begin{aligned} k(\bar{x}^i, \bar{x}^j) &= \exp \left[\frac{-(\bar{x}^i - \bar{x}^j)^2}{2\sigma^2} \right] \\ &= \exp \left[\frac{-\{(\bar{x}^{iT} \cdot \bar{x}^i) + (\bar{x}^{jT} \cdot \bar{x}^j) - 2(\bar{x}^{iT} \cdot \bar{x}^j)\}}{2\sigma^2} \right] \\ &= \exp \left[-\frac{\bar{x}^{iT} \cdot \bar{x}^i}{2\sigma^2} \right] \cdot \exp \left[-\frac{\bar{x}^{jT} \cdot \bar{x}^j}{2\sigma^2} \right] \cdot \exp \left[\frac{\bar{x}^{iT} \cdot \bar{x}^j}{\sigma^2} \right] \end{aligned}$$

Terlihat bahwa dari fungsi yang diberikan, titik i dan j berdimensi n terpetakan ke sebuah koordinat baru dari fungsi $k(\bar{x}^i, \bar{x}^j)$. Semua kombinasi titik dari sebuah matriks A dapat digambarkan sebagai berikut:

$$K = \begin{bmatrix} k(\bar{x}^1, \bar{x}^1) & \dots & k(\bar{x}^1, \bar{x}^n) \\ \vdots & \ddots & \vdots \\ k(\bar{x}^n, \bar{x}^1) & \dots & k(\bar{x}^n, \bar{x}^n) \end{bmatrix}$$

yang mana matriks ini bisa dihasilkan dalam bentuk *inner product*.

Principal Component Analysis

Principal Component Analysis adalah analisis yang mentransformasikan variable lama menjadi variable baru dengan mereduksi sejumlah variabel tanpa harus kehilangan sebagian besar variabel aslinya. Algoritma dari PCA dapat dijelaskan sebagai berikut:

1. Mencari rata-rata untuk setiap variabel, lalu menghitung kovarian
2. Mencari nilai eigen dan vector eigen dari matriks kovariansi empiric yang diperoleh
3. Menghitung proporsi variansi dari masing-masing PC beserta nilai akumulasi untuk q-PC pertama

Autoregressive

Autoregresi merupakan bentuk model regresi dengan mengamati nilai sekarang berdasarkan nilai-nilai sebelumnya dengan selang waktu tertentu. Peramalan autoregresi dapat ditulis sebagai berikut:

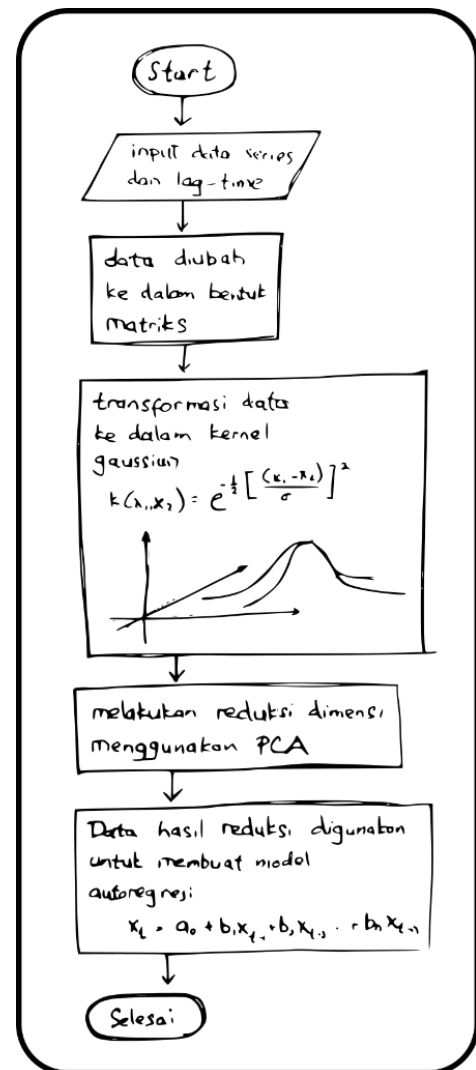
$$x_t = b_0 + b_1x_{(t-1)} + \dots + b_nx_{(t-n)}$$

$$\text{atau} \\ x_t = b_0 + \sum_{i=1}^n b_nx_{(t-n)}$$

Proses yang digunakan umumnya adalah regresi linear yang variable independennya adalah data pada waktu sebelumnya.

METODE PENELITIAN

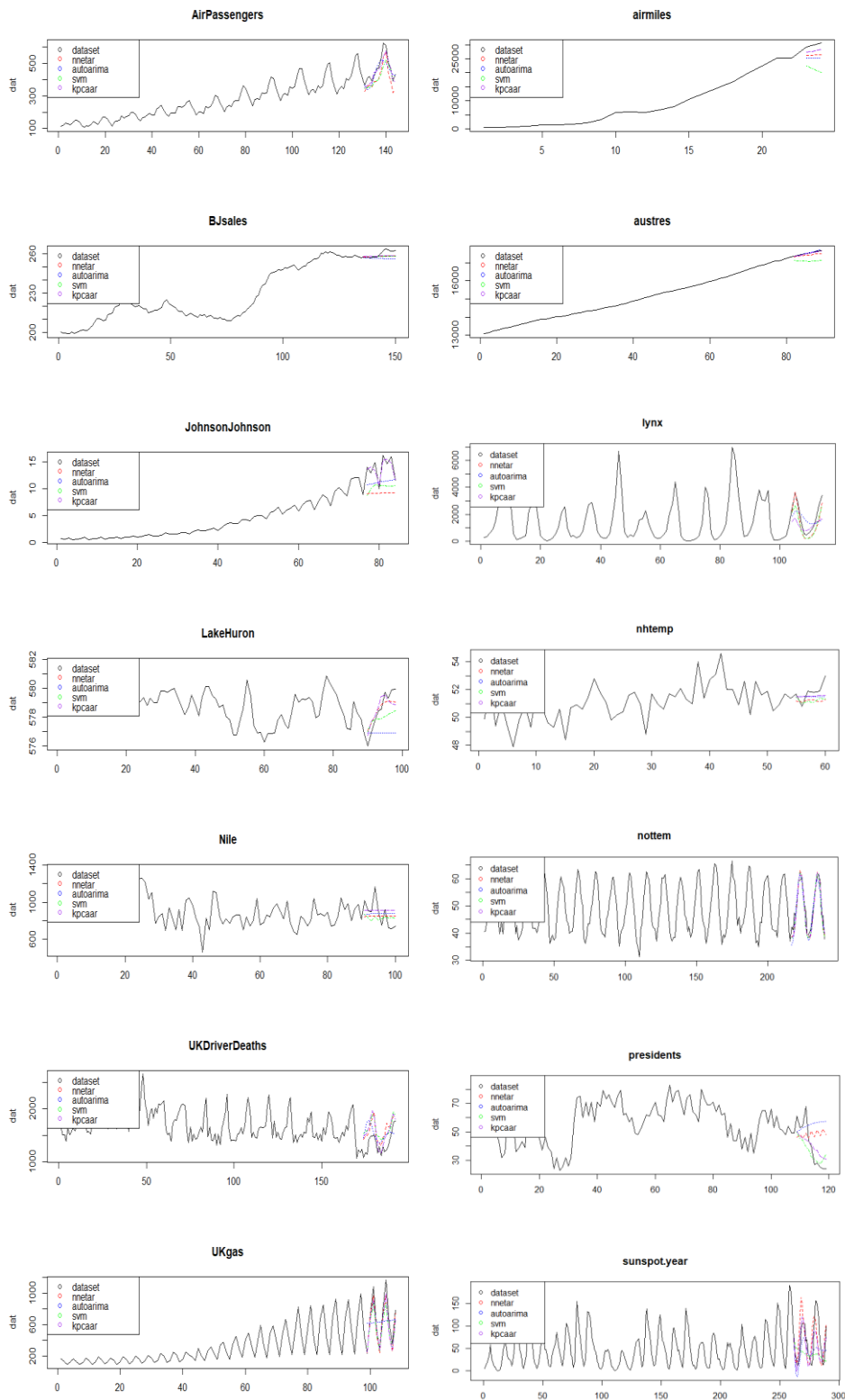
Alur kerja KPCA-AR terlihat seperti pada gambar 1. Algoritma KPCA-AR ini dilakukan dengan beberapa tahapan. Tahapan pertama, data yang akan dilakukan peramalan diubah ke dalam bentuk matriks dengan jumlah kolom sebanyak n-lag waktu. Pada peramalan autoregressive, data yang telah menjadi matriks ini kemudian dibuat sebuah model dengan metode regresi linier. Perbedaannya pada peramalan ini adalah data di transformasi terlebih dahulu ke dalam bentuk kernel gaussian.



Gambar 1. Diagram Kerangka Pikir Penelitian

Setelah data di transformasi ke kernel gaussian, dilakukan reduksi dimensi menggunakan PCA. Dilakukannya reduksi dimensi pada data yang telah di transformasi bertujuan untuk mengurangi beban komputasi pada perhitungan tanpa mengurangi nilai dari data yang ada secara signifikan.

Selanjutnya dilakukan pengujian metode KPCA-AR. Pengujian dilakukan dengan menggunakan beberapa data (<http://www.eio.upc.edu/~pau/cms/rdata/datasets.html>). Pengujian dilakukan dengan membandingkan akurasi hasil peramalan KPCAAR dengan beberapa metode yang umum digunakan seperti SVM, ANN, dan ARIMA. Parameter KPCAAR seperti lag_time dan sigma dipilih secara manual, dilakukan beberapa kali percobaan, lalu



Gambar 6. Grafik Peramalan Masing-Masing Metode

dipilih hasil yang terbaik untuk dilakukan perbandingan.

HASIL DAN PEMBAHASAN

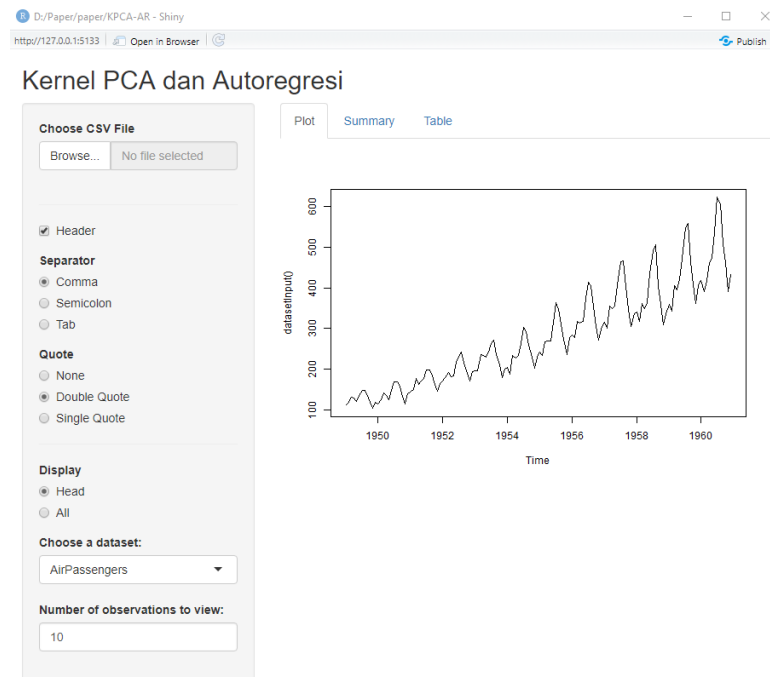
Algoritma KPCAAR dibuat dengan menggunakan Bahasa pemrograman R.

Untuk ANN dan ARIMA, menggunakan function “nnetar” dan “auto.arima” dari library “Forecast”. Sedangkan untuk SVM menggunakan function “svm” dari library “e1071”.

Dari percobaan data yang dilakukan, diperoleh hasil seperti pada Tabel 1 dan

Tabel 1. MAPE untuk Setiap Metode

Dataset	nnetar	autoarima	svm	kpcaar	lag_time
AirPassengers	15.94372321	7.441935163	10.7463359	7.864531784	12
BJsales	0.862745266	1.322658056	0.923395793	0.84982028	12
JohnsonJohnson	31.6261733	19.62258008	33.73841313	5.762246031	12
LakeHuron	0.103826264	0.312135633	0.174800607	0.107171035	3
Nile	13.77403041	13.6604378	13.74082359	13.46892396	6
UKDriverDeaths	18.63374415	18.45316595	20.98493376	18.1515543	12
Ukgas	12.36497035	53.62159405	15.59796615	12.10370445	6
airmiles	11.94842902	15.16495802	28.74914311	6.716193478	3
austres	0.523319045	0.192920899	2.373558751	0.097649517	6
lynx	33.0095407	70.94256864	39.05565862	42.13972704	9
nhtemp	1.451560646	1.13602354	1.299848029	1.043626014	3
nottem	4.126216051	4.928378424	4.02801591	3.707257135	6
presidents	51.78385659	65.73650925	20.26320618	22.51766703	12
sunspot.year	88.79173087	68.53967663	73.37327652	72.87463634	7



Gambar 3. Aplikasi KPCAAR menggunakan R

Gambar 2. Dari tabel 1, 9 dari 14 dataset yang diberikan menunjukkan bahwa KPCAAR dapat memberikan nilai MAPE yang lebih kecil dari metode peramalan lainnya.

Pada Gambar 2 dan Tabel 1 terlihat bahwa peramalan yang dihasilkan oleh KPCAAR memberikan nilai MAPE yang lebih kecil dibandingkan dengan yang lainnya, diantaranya yaitu BJsales, JohnsonJohnson, Nile, UKDriverDeaths, Ukgas, airmiles, austres, nhtemp, dan nottem.

KESIMPULAN DAN SARAN

KPCAAR berhasil dibuat menggunakan bahasa pemrograman R, sedangkan untuk algoritma lainnya menggunakan library “forecast” dan “e1071”. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa KPCAAR secara umum mampu memberikan hasil peramalan yang baik, terlihat dari MAPE yang dihasilkan lebih kecil dari beberapa metode lainnya yang digunakan untuk dataset yang

sudah diberikan. Hasil juga dipengaruhi oleh lag_time dan parameter sigma yang diberikan. Oleh karena itu, perlu adanya penelitian lanjutan untuk mengetahui nilai lag_time dan sigma yang optimal agar memberikan hasil yang semakin baik.

DAFTAR PUSTAKA

- [1] A. Luqman, A. Kusworo, W. Achmad. *Support Vector Machine untuk Prakiraan Beban Listrik Harian pada Sektor Industri, Jurnal Teknologi Elektro dan Kejuruan*, vol. 24, no. 2, 2015, pp. 23-28.
- [2] L. Tri. *Perbandingan Algoritma Neural Network (NN) dan Support Vector Machines (SVM) dalam Peramalan Penduduk Miskin di Indonesia, Jurnal Online ICT STMIK IKMI*, vol. 10, no. 1, 2013, pp. 49-57.
- [3] R. Santosha, K. Naveena, S. Subedar, S. Abhishek. *Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India, International Journal of Current Microbiology and Applied Sciences*, vol. 6, 2017, pp. 1721 – 1726.
- [4] M. Sung-Hwan, L. Jumin, H. Ingoo. *Hybrid genetic algorithm and support vector machines for bankruptcy prediction, Expert Systems with Applications*, vol. 3, 2006, pp. 652-660.