

FUNGSI KERNEL UNTUK ESTIMASI PERTUMBUHAN PRODUKSI INDUSTRI MIKRO/KECIL DI DAERAH ISTIMEWA YOGYAKARTA

Suparna Parwodiwiyo

Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta

Email: parno987@gmail.com

Abstract. *The contribution of micro and small industries (IMK) in the Special Region of Yogyakarta is far greater than the national average, both in terms of value added and employment. At present the production growth rate is needed in a faster period and requires estimation. This study aims to use a kernel estimator with five kernel functions to estimate the growth rate of IMK production in the Special Region of Yogyakarta. Analysis with the application of R shows that by using the minimum mean square error criteria, five functions produce various optimal bandwidth values but have a similar estimated curve. The Gauss-normal kernel function produces the smallest MSE value of 10,38 and gets the estimated value of the growth of IMK production in the next quarter of -0,38 percent.*

Keywords: *Bandwidth, Kernel Functions, MSE, IMK*

Abstrak. Kontribusi industri mikro dan kecil (IMK) di Daerah Istimewa Yogyakarta jauh lebih besar dibandingkan angka rata-rata nasional, baik dari sisi nilai tambah maupun penyerapan tenaga kerja. Saat ini angka pertumbuhan produksi diperlukan dalam perodesasi yang lebih cepat dan memerlukan estimasi. Penelitian ini bertujuan menggunakan estimator kernel dengan lima fungsi kernel untuk melakukan estimasi angka pertumbuhan produksi IMK di Daerah Istimewa Yogyakarta. Analisis dengan aplikasi R menunjukkan bahwa dengan menggunakan kriteria *mean square error* minimum, lima fungsi menghasilkan berbagai nilai bandwidth optimal tetapi memiliki estimasi kurva yang serupa. Fungsi kernel Gauss-normal menghasilkan nilai MSE terkecil yaitu sebesar 10,38 dan mendapatkan hasil estimasi nilai pertumbuhan produksi IMK pada triwulan berikutnya sebesar -0,38 persen.

Kata kunci: *Bandwith, Fungsi Kernel, MSE, IMK*

1. Pendahuluan

Kontribusi sektor industri manufaktur dalam pembangunan ekonomi nasional cukup besar. Kontribusi ini dapat dilihat dari nilai tambah industri manufaktur terhadap Produk Domestik Bruto (PDB). Kontribusi industri manufaktur pada tahun 2019 sebesar 19,7 persen. Dengan demikian pertumbuhan produksi yang berkelanjutan disertai dengan pertumbuhan jumlah usaha dapat memberikan dampak positif bagi perekonomian nasional. Sebab, industri manufaktur diyakini menawarkan peluang tingkat pekerjaan dan pendapatan yang lebih baik, sehingga pengembangan sektor ini dapat dijadikan sebagai alat untuk melakukan transformasi struktural di Indonesia. Hal ini diperkuat bahwa industri manufaktur memberikan tingkat produktivitas yang lebih tinggi dibandingkan dengan sektor lain, khususnya pertanian[4].

Salah satu persoalan yang menjadi perhatian banyak pihak adalah informasi pertumbuhan produksi dengan perodesasi yang cepat. Ketersediaan data dan informasi yang akurat dan cepat sangat dibutuhkan oleh pemerintah sebagai dasar pengambilan kebijakan di sektor industri manufaktur, baik industri skala besar-sedang (IBS) atau skala industri mikro-kecil (IMK). Saat ini penyediaan data pertumbuhan produksi secara triwulanan tetapi masih ditemukan kendala dalam tingkat respon perusahaan dan kualitas isian, serta hasil yang sangat berfluktuatif. Ada beberapa metode untuk mendapatkan angka pertumbuhan produksi secara lebih cepat untuk mengatasi hal tersebut, salah satunya dengan cara estimasi.

Penggunaan metode nonparametrik tepat digunakan pada beberapa jenis data dan bebas dari tuntutan asumsi data, salah satunya yaitu data *time series*[15]. Data *time series* sering bersifat fluktuatif serta galatnya diasumsikan saling berkorelasi sehingga perlu analisis tersendiri[12]. Pendekatan nonparametrik yang tepat digunakan adalah pendekatan dengan

menggunakan estimator fungsi kernel. Estimator kernel paling tidak memiliki lima macam fungsi yaitu Gauss (normal), Epanechnikov, Uniform(box), Biweight, dan Triweight[5][6][10].

Untuk kasus Daerah Istimewa Yogyakarta peran IMK jauh lebih besar dibandingkan angka rata-rata nasional, baik dari sisi nilai tambah maupun penyerapan tenaga kerja (28%). Akan tetapi angka pertumbuhan per triwulannya sangat bervariasi sehingga perlu perhatian yang lebih tinggi. Berdasarkan hal tersebut, maka menarik untuk melakukan analisis dengan menggunakan estimator kernel pada data pertumbuhan produksi industri pengolahan/manufaktur khusus IMK secara triwulanan di Daerah Istimewa Yogyakarta.

2. Metode

Sumber data yang digunakan adalah data sekunder, yaitu data pertumbuhan produksi IMK yaitu usaha industri manufaktur dengan tenaga kerja kurang dari 20 orang. Data tersedia setiap triwulan tahun 2015-2019 yang diambil dari Berita Resmi Statistik BPS Provinsi Daerah Istimewa Yogyakarta pada laman <http://www.yogyakarta.bps.go.id>. Estimasi kurva regresi dilakukan dengan menggunakan estimator kernel, dengan lima fungsi kernel yaitu Uniform, Epanechnikov, Biweight, Triweight, dan Gauss-normal dengan menggunakan bantuan software R i386 versi 3.6.1.

Regresi parametrik biasanya membutuhkan asumsi pola data tertentu, misalnya data berdistribusi normal, tidak ada autokorelasi, dan lain-lain. Regresi nonparametrik biasanya jadi jalan keluar untuk data yang tidak dapat memenuhi asumsi data tersebut⁷. Penelitian sebelumnya yang memodelkan data nilai kurs rupiah terhadap dollar Amerika Serikat pada bulan Januari 2012 dengan menggunakan penalized spline filter[15]. Analisis data inflasi di Indonesia menggunakan model regresi polinomial lokal dengan kriteria generalized cross validation (GCV) minimum untuk menentukan titik lokal optimal[13].

Misalkan untuk n pengamatan yang independen, (x_i, y_i) dengan $i=1,2, \dots, n$ maka model regresi nonparametrik secara umum dapat ditulis dengan:

$$y_i = f(x_i) + \varepsilon_i, \quad i=1,2,\dots, n \quad (1)$$

y_i adalah variabel terikat ke- i , $f(x_i)$ adalah fungsi regresi yang tidak diketahui bentuk kurva regresinya dan ε_i adalah random error atau galat acak yang diasumsikan independen dan identik dengan rata-rata 0 dan ragam σ^2 [1].

Salah satu teknik dalam regresi nonparametrik adalah estimator fungsi kernel. Dengan memberi bobot pada setiap pengamatan, dengan memilih fungsi K , sehingga pengamatan yang lebih dekat ke x akan memberi sumbangan yang lebih besar terhadap $f(x)$. Fungsi K ini merupakan fungsi pembobot yang dinamakan fungsi kernel. Bentuk bobot kernel ditentukan oleh fungsi kernel K , sedangkan ukuran bobotnya ditentukan oleh parameter pemulus h yang disebut *bandwidth* [14].

Secara umum kernel K dengan bandwidth h didefinisikan sebagai:

$$K_h(u) = K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right) \quad (2)$$

Untuk $-\infty < h < \infty$; serta memenuhi:

- (a) $K(u) \geq 0$,
- (b) $\int_{-\infty}^{\infty} K(u) d(u) = 1$,
- (c) $\int_{-\infty}^{\infty} u^2 K(u) d(u) = \sigma^2 > 0$,
- (d) $\int_{-\infty}^{\infty} u K(u) d(u) = 0$.

Adapun beberapa jenis fungsi kernel yang tersedia pada Modul R dapat dilihat dalam Tabel 1. menurut jenisnya [8].

Tabel 1. Jenis-jenis Fungsi Kernel

No	Fungsi Kernel	Bentuk Fungsi
1	Uniform	$K(u) = \frac{1}{2}; I(u \leq 1)$
2	Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2) ; I(u \leq 1)$
3	Biweight	$K(u) = \frac{15}{16}(1 - u^2)^2 ; I(u \leq 1)$
4	Triweight	$K(u) = \frac{35}{32}(1 - u^2)^3 ; I(u \leq 1)$
5	Gauss-normal	$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(-u^2)\right)$

dengan I adalah fungsi indikator.

Regresi kernel adalah salah satu teknik statistika nonparametrik yang dapat digunakan untuk menduga fungsi regresi $f(x_i)$ pada model regresi nonparametrik pada persamaan (1). Pada tahun 1964, Nadaraya dan Watson mengusulkan penduga/estimator kernel dalam regresi nonparametrik yang kemudian sering disebut dengan estimator Nadaraya-Watson[10] yaitu:

$$\hat{f}(x) = \frac{\sum_{i=1}^n K_h(x-x_i)y_i}{\sum_{i=1}^n K_h(x-x_i)} \tag{3}$$

$$= \sum_{i=1}^n W_{hi}(x)y_i \tag{4}$$

dengan $W_{hi}(x) = \frac{K(\frac{x-x_i}{h})}{\sum_{k=1}^n K(\frac{x-x_k}{h})}$ (5)

Selanjutnya, untuk memperoleh nilai duga masing-masing pengamatan yaitu x_i , maka diperoleh:

$$\hat{f}(x_i) = \sum_{j=1}^n W_{hj}(x_i)y_j \tag{6}$$

dengan $W_{hj}(x) = \frac{K(\frac{x_i-x_j}{h})}{\sum_{k=1}^n K(\frac{x_i-x_k}{h})}$ $i,j=1,2, \dots, n$ (7)

Pemilihan *bandwidth* jauh lebih penting dibandingkan dengan memilih fungsi kernel[3]. *Bandwidth* (h) adalah parameter pemulus (*smoothing*) yang berfungsi untuk mengontrol kemulusan dari kurva yang diduga. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *undersmoothing* yaitu kurva yang sangat kasar dan sangat fluktuatif, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *oversmoothing* yaitu kurva yang sangat mulus, tetapi tidak sesuai dengan pola data. Oleh karena itu, perlu dipilih *bandwidth* yang optimal untuk menghasilkan kurva yang optimal.

Suatu kriteria untuk h akan dibatasi pada kelas penduga linier, dimana untuk setiap h ada matriks H(h) berukuran n x n, H(h) simetris dan semidefinit positif, sehingga $fh=H(h)Y$ dengan elemen-elemen H(h) adalah H_{ij} yang diperoleh dari menuliskan kembali persamaan (7) sebagai:

$$H_{ij} = \frac{K(\frac{x_i-x_j}{h})}{\sum_{k=1}^n K(\frac{x_i-x_k}{h})} \quad i,j=1,2, \dots, n \tag{8}$$

[5].

Salah satu metode untuk mendapatkan h optimal adalah dengan menggunakan kriteria *generalized cross validation* (GCV) atau MSE (*mean square error*) yang didefinisikan seperti pada persamaan 9 dan 10.

$$GCV(h) = \frac{MSE}{\left(\frac{1}{n}tr(I-H(h))\right)^2} \quad (9)$$

dengan

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - f_h(X_i))^2 \quad (10)$$

I menyatakan matriks identitas dan tr menyatakan *trace*/teras dari matriks $(I-H(h))$. Selain digunakan dalam penentuan nilai GCV, MSE merupakan salah satu metode yang dapat digunakan untuk mengukur ketepatan suatu estimator. Semakin kecil nilai MSE yang dihasilkan, maka semakin baik hasil estimasinya[2].

Regresi nonparametrik dengan fungsi kernel pada data deret waktu menggunakan tiga pemodelan matematika sebagai berikut:

1. Model (S) yaitu suatu barisan stasioner $\{(X_i, Y_i), i=1,2,\dots, n\}$ (diperbolehkan dependen stokastik) telah diobservasi dan berdasarkan hasil observasi akan diestimasi:

$$f(x)=E(Y/X=x) \quad (11)$$

2. Model (T) yaitu suatu runtun waktu $\{Z_i, i \geq 1\}$ telah diobservasi dan akan diprediksi Z_{n+1} dengan:

$$f(x)=E(Z_{n+i}/Z_n=x) \quad (12)$$

3. Model (C) yaitu error observasi $\{e_m\}$ dalam model regresi dengan rancangan tetap membentuk barisan variabel random yang berkorelasi.

$$Y_{in} = f\left(\frac{i}{n}\right) + e_{in} \quad (13)$$

Permasalahan model (T) dapat dipetakan dengan permasalahan model (S) dengan mendefinisikan dalam runtun waktu $\{Z_i, i \geq 1\}$, yaitu nilai lag Z_{i-1} sebagai X_i dan Z_i sebagai Y_i . Selanjutnya, masalah prediksi Z_{n+1} dari $\{Z_i, i=1,2,\dots, n\}$ dapat dipandang sebagai masalah pemulusan regresi untuk

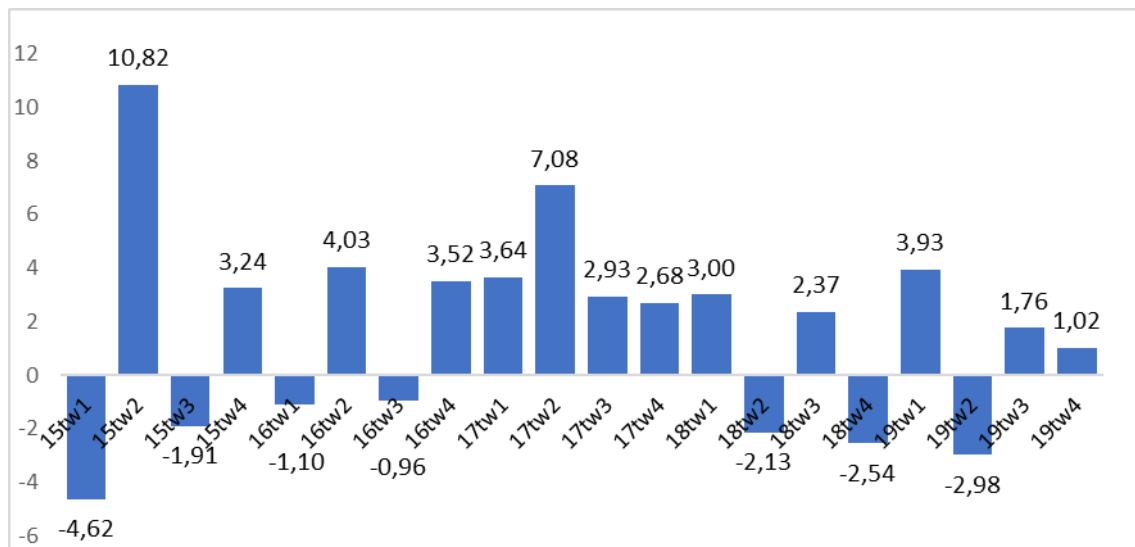
$$\{(X_i, Y_i)\} = \{(Z_{i-1}, Z_i)\}; i=1,2,\dots, n \quad (14)$$

[3][7].

3. Hasil dan Pembahasan

Gambar 1. memperlihatkan bahwa data pertumbuhan produksi triwulanan industri pengolahan/manufaktur kategori IMK di Daerah Istimewa Yogyakarta pada periode 2015-2019 tersebut tidak membentuk pola linier tertentu. Selama periode data 20 triwulan tersebut nilai pertumbuhan produksi industri pengolahan IMK cenderung mengalami fluktuasi. Angka pertumbuhan produksi IMK di Daerah Istimewa Yogyakarta tertinggi terjadi pada triwulan 2 tahun 2015 yang mencapai 10,82 persen. Sementara angka pertumbuhan produksi IMK terendah terjadi pada triwulan 1 tahun 2015 yang sebesar -4,62 persen. Seperti sudah dibahas sebelumnya bahwa penggunaan analisis regresi linier tidak tepat diterapkan sebagai model analisis karena keterbatasan data dan sifatnya yang fluktuatif. Berdasarkan keterbatasan jumlah data dan pola data tersebut, maka data pertumbuhan produksi triwulanan IMK perlu diestimasi dengan menggunakan pendekatan estimator fungsi kernel.

Berdasarkan metode analisis yang dipilih langkah pertama untuk melakukan estimasi dengan menggunakan estimator fungsi kernel adalah menentukan nilai *bandwidth* optimal masing-masing fungsi kernel yang digunakan yaitu Uniform, Epanechnikov, Biweight, Triweight, dan Gauss-normal. Hal ini dilakukan karena fungsi ini sudah tercakup dalam prosedur KernSmooth pada aplikasi R yang digunakan untuk proses pengolahan data. Nilai *bandwidth* optimal diperoleh berdasarkan kriteria nilai MSE yang minimum.



Gambar 1. Perkembangan Angka Pertumbuhan Produksi IMK di Daerah Istimewa Yogyakarta

Secara ringkas nilai *bandwidth* optimal dari kelima fungsi kernel disajikan pada Tabel 2. Tabel 2 menunjukkan nilai *bandwidth* optimal dari masing-masing fungsi kernel dengan nilai MSE minimumnya. Tabel tersebut memperlihatkan bahwa masing-masing fungsi kernel memiliki nilai *bandwidth* optimal yang bervariasi.

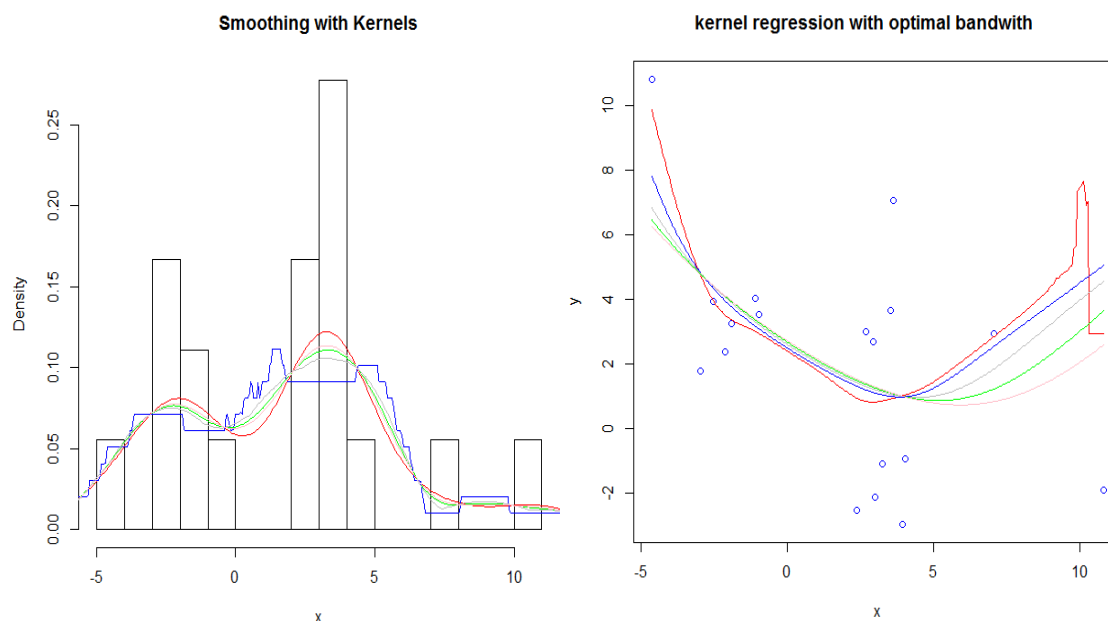
Tabel 2. Nilai *Bandwidth* Optimal Fungsi Kernel

Nomor	Fungsi kernel	Bandwith Optimal	Nilai SSE	Nilai MSE	Nilai Yduga TW1 2020
1	Uniform/box	2,717	254,82	14,16	0,25
2	Epanechnikov	3,457	233,42	12,97	0,75
3	Biweight	4,095	230,31	12,80	0,91
4	Triweight	4,650	228,15	12,68	1,02
5	Gauss-normal	1,562	186,86	10,38	-0,38

Sumber: data diolah (2020)

Selanjutnya, nilai *bandwidth* optimal tersebut digunakan untuk memperoleh hasil estimasi kurva regresi kelima fungsi *kernel*. Adapun hasil estimasi kurva kelima fungsi *kernel* disajikan pada Gambar 2. Gambar 2 menunjukkan bahwa kelima fungsi *kernel* dengan masing-masing nilai *bandwidth* optimalnya menghasilkan kurva yang hampir mirip.

Tabel 2. juga memperlihatkan nilai MSE yang dihasilkan oleh kelima fungsi kernel. Nilai MSE menunjukkan seberapa bagus hasil estimasi yang dihasilkan oleh estimator kernel dengan menggunakan fungsi kernel yaitu Uniform, Epanechnikov, Biweight, Triweight, dan Gauss-normal. Semakin kecil nilai MSE yang dihasilkan, berarti semakin bagus hasil estimasinya. Dari hasil analisis didapatkan bahwa fungsi kernel dengan pendekatan Gauss-normal dengan *bandwith* optimal sebesar 1,562 mendapatkan nilai MSE minimal yaitu sebesar 10,38. Nilai MSE terbesar yaitu 14,16 didapatkan dari model dengan pendekatan Uniform/box.



Gambar 2. Plot Data dan Kurva Hasil Estimasi Kelima Fungsi Kernel

Kelima fungsi kernel menghasilkan MSE yang nilainya tidak jauh berbeda antara satu sama lain sehingga dari nilai MSE tersebut dapat memperkuat bahwa kelima fungsi kernel menghasilkan estimasi kurva yang mirip. Hal ini mendukung pendapat penelitian sebelumnya yang menyatakan bahwa pemilihan *bandwidth* jauh lebih penting dibandingkan pemilihan fungsi kernel dikarenakan telah terbukti pemilihan fungsi kernel yang berbeda akan menghasilkan estimasi kurva yang hampir sama[3][5].

Model fungsi kernel terbaik dengan nilai MSE terkecil dapat digunakan untuk memprediksikan nilai Y yang akan datang. Dengan menggunakan pemodelan kernel Gauss-normal berdasarkan data Triwulan I 2015 – Triwulan IV 2019, didapatkan hasil prediksi pertumbuhan produksi IMK di Daerah Istimewa Yogyakarta pada Triwulan I 2020 mengalami sedikit penurunan namun tidak signifikan. Didapatkan angka pertumbuhan produksi IMK pada Triwulan I 2020 sebesar -0,38 persen. Hanya saja disayangkan bahwa data pertumbuhan produksi IMK Triwulan I 2020 data dari BPS tidak terbit, sehingga tidak dapat dibandingkan angka estimasi dengan angka hasil survei.

4. Kesimpulan

Pendekatan estimator fungsi kernel yaitu menggunakan fungsi kernel Uniform/box, Epanechnikov, Biweight, Triweight, dan Gauss-normal pada data pertumbuhan produksi IMK menghasilkan nilai *bandwidth* optimal yang bervariasi. Namun, dengan menggunakan nilai *bandwidth* optimal masing-masing menghasilkan kurva regresi yang berimpit dan mengikuti pola data aktual. Fungsi kernel Gauss-normal menghasilkan nilai MSE terkecil yaitu sebesar 10,38 dan mendapatkan hasil estimasi nilai pertumbuhan produksi IMK pada triwulan berikutnya sebesar -0,38 persen.

Dilihat berdasarkan nilai MSE yang dihasilkan, masing-masing fungsi memiliki nilai MSE yang tidak jauh berbeda antara satu sama lain sehingga memperkuat bahwa kelima fungsi menghasilkan estimasi kurva yang hampir sama. Jadi, dapat dikatakan bahwa penerapan estimator kernel pada data time series juga menghasilkan kesimpulan bahwa pemilihan *bandwidth* jauh lebih penting dibandingkan dengan pemilihan fungsi kernel. Meskipun

demikian kita tetap bisa memilih fungsi kernel yang memiliki nilai MSE terkecil untuk mendapatkan estimasi yang lebih tepat.

Ucapan Terima Kasih

Terima kasih disampaikan kepada Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta atas perkenan datanya untuk dilakukan pengolahan dan dibahas lebih lanjut.

Daftar Pustaka

- [1] Adamiak K, Kim H, Slot K, Kim H, Slot K, Projec A..2020. Accelerating Projections to Kernel-Induced Spaces by Feature Approximation. *Pattern Recognit Lett J homepage www.elsevier.com Accel*. Published online 2020:1-10. doi:10.1016/j.patrec.2020.05.029
- [2] Aydin, D. A. 2007. Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel regression. *Int J Math Comput Sci Eng*. 1(12):416-420. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.192.8167&rep=rep1&type=pdf>
- [3] Barry, D, Hardle W. 1994. *Applied Nonparametric Regression*. Vol 156. doi:10.2307/2982873
- [4] BPS DIY. 2020. *Daerah Istimewa Yogyakarta Dalam Angka 2020*. Yogyakarta.
- [5] Chen, YC. 2018. Modal regression using kernel density estimation: A review. *Wiley Interdiscip Rev Comput Stat*. 10(4):1-29. doi:10.1002/wics.1431
- [6] Čížek, P, Sadikoğlu S. 2020. Robust nonparametric regression: A review. *Wiley Interdiscip Rev Comput Stat*. 12:1-16. doi:10.1002/wics.1492
- [7] Elisa, MM. 20019. Membandingkan Regresi 4PL dan Linier Fit Untuk Verivikasi Hormon 17β -Estradiol. 5(1).
- [8] Laome, L. 2010. Perbandingan model regresi nonparametrik dengan regresi spline dan kernel. *Jimt*. 7(1):1-7. doi:10.22487/j2450766X.2010.v7.i1.28
- [9] Lestari, B, Chamidah N, Saifudin T. 2019. Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. *J Mat Stat dan Komputasi*. 15(2):20-25. doi:10.20956/jmsk.v15i2.5710
- [10] Mahmoud, HFF. 2019. *Parametric versus Semi and Nonparametric Regression Models..* <http://arxiv.org/abs/1906.10221>
- [11] Milton, P, Coupland H, Giorgi E, Bhatt S. 2019. Spatial analysis made easy with linear regression and kernels. *Epidemics*. 29(August):100362. doi:10.1016/j.epidem.2019.100362
- [12] Santoso, R, Prahutama, A, Devi, AR. 2018. Analisis Data Inflasi Indonesia Menggunakan Metode Fourier dan Wavelet Multiscale Autoregressive. In: *Seminar Nasional Variansi*. 185-196.
- [13] Suparti, Prahutama, A. 2016. Pemodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal Pada Beban Listrik Di Kota Semarang. *Media Stat*. 9(2):85-93. doi:10.14710/medstat.9.2.85-93
- [14] Wandresen, RR, Pellicico S, Koehler HS, Sanquerta CR, Behling A. 2019. Nonparametric Method:Kernel Density Estimation Aplied to Forestry Data. *Floresta*. 49(3):561-570. doi:10.5380/rf.v49
- [15] Wuleng, A, Islamiyati A, Herdiani E. 2014. Pemodelan Data Time Series Dengan Penalized Spline Filter. *UNDIP*. 1(1):100362. doi:10.1016/j.epidem.2019.100362