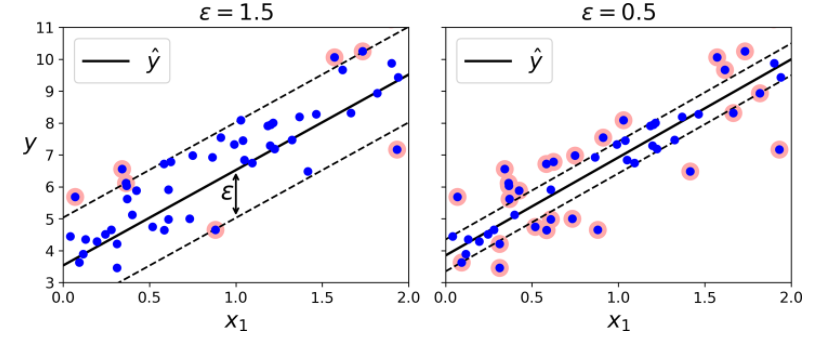
Seperti yang telah disebutkan di awal modul, selain bisa menyelesaikan masalah klasifikasi, support vector juga bisa dipakai untuk prediksi data kontinu yaitu kasus regresi.

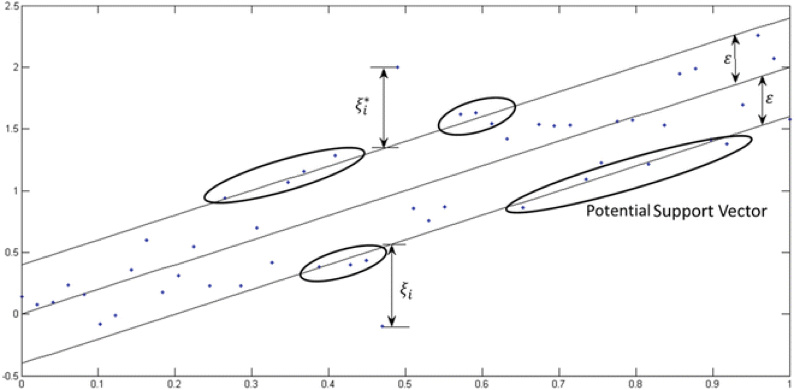
Support Vector Regression (SVR) menggunakan prinsip yang sama dengan SVM pada kasus klasifikasi. Perbedaannya adalah jika pada kasus klasifikasi, SVM berusaha mencari ‘jalan’ terbesar yang bisa memisahkan sampel-sampel dari kelas berbeda, maka pada kasus regresi SVR berusaha mencari jalan yang dapat menampung sebanyak mungkin sampel di ‘jalan’.

Perhatikan gambar berikut untuk melihat contoh SVR [4].

[](https://www.dicoding.com/academies/184/tutorials/8457?from=8452)

Seperti dijelaskan oleh Garon dalam Hands-On Machine Learning with Scikit Learn [4], gambar di atas menunjukkan dua model Regresi SVM linier yang dilatih pada beberapa data linier acak, satu dengan margin besar (ϵ = 1,5) dan yang lainnya dengan margin kecil (ϵ = 0,5). Lebar jalan dikontrol oleh hyperparameter ϵ, yang juga disebut maksimum eror. Menambahkan data training ke dalam margin tidak akan mempengaruhi prediksi model. Oleh karena itu, model disebut sebagai ϵ-insensitivity (tidak sensitif-ϵ).

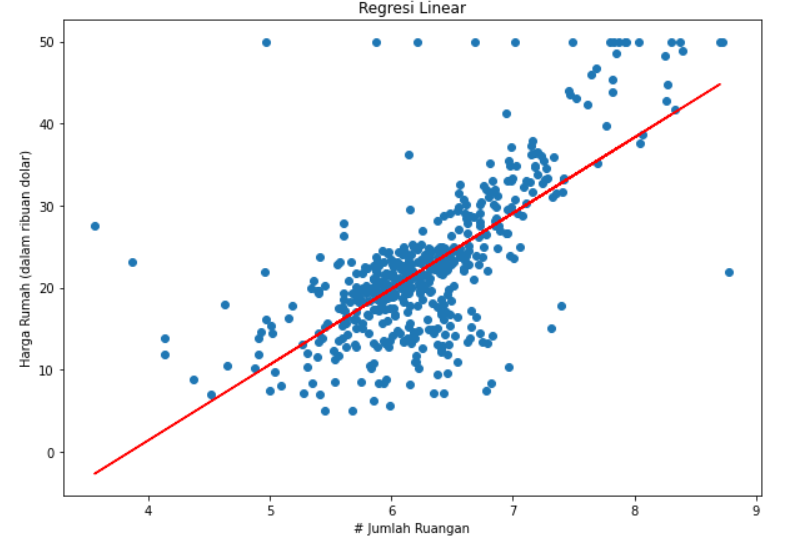
Berbeda dengan SVM di mana support vector adalah 2 sampel dari 2 kelas berbeda yang memiliki jarak paling dekat, pada SVR support vector adalah sampel yang menjadi pembatas jalan yang dapat menampung seluruh sampel pada data. M. Awad dan R. Khanna dalam bab 4 bukunya [[14]](https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4) mengilustrasikan support vector pada SVR sebagai berikut.

[](https://www.dicoding.com/academies/184/tutorials/8457?from=8452)

Mari kita ambil contoh implementasi SVR pada kasus prediksi harga rumah di kota Boston, Amerika Serikat, menggunakan dataset yang sangat populer untuk kasus regresi: “Boston Housing Price.csv”. Pertama, kita akan melihat bagaimana hasil prediksi regresi linear sederhana pada data ini, kemudian kita akan membandingkan hasilnya dengan SVR.

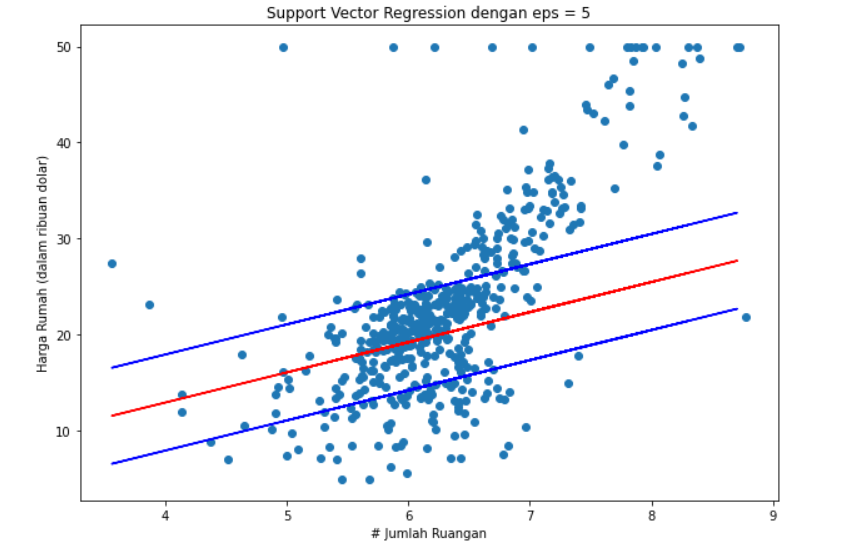
Sedikit mengingat kembali tentang regresi linear yang telah diulas di modul sebelumnya, ukuran performa untuk permasalahan regresi linear adalah Root Mean Square Error (RMSE). RMSE memberi gambaran tentang seberapa banyak kesalahan dalam prediksi yang dibuat oleh sistem. Tujuannya tentu saja untuk mendapatkan eror atau tingkat kesalahan seminimal mungkin.

Pada kasus prediksi harga rumah di Boston, regresi linear akan memberikan hasil plot sebagai berikut.

[](https://www.dicoding.com/academies/184/tutorials/8457?from=8452)

Mari kita coba implementasikan SVR pada dataset yang sama. Salah satu kelebihan SVR dibanding regresi linear adalah SVR memberi kita fleksibilitas untuk menentukan seberapa banyak kesalahan yang dapat diterima dalam model kita. Algoritma SVR akan menemukan garis yang cocok (hyperplane) agar sesuai dengan data. Kita bisa mengatur parameter ϵ untuk mendapatkan akurasi model yang kita inginkan.

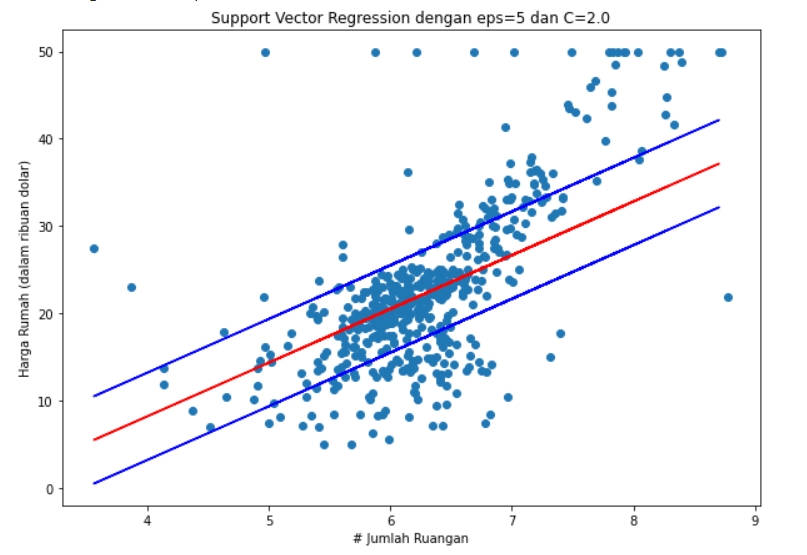
Jika kita pilih nilai ϵ = 5 dan kita plot datanya, hasilnya adalah sebagai berikut.

[](https://www.dicoding.com/academies/184/tutorials/8457?from=8452)

Garis merah pada gambar menunjukkan garis regresi, sedangkan garis biru menunjukkan margin dari eror, ϵ, yang telah kita atur nilainya tadi dengan ϵ = 5 (atau dalam skala ribuan berarti senilai $ 5,000).

Dari gambar di atas Anda mungkin bisa langsung melihat bahwa algoritma SVR ini tidak bisa memberikan hasil prediksi yang baik untuk seluruh data sebab beberapa titik masih berada di luar batas. Oleh karena itu, kita perlu menambahkan parameter lain pada algoritma yaitu parameter C yang disebut sebagai regularization parameter atau parameter keteraturan. Ada juga yang menyebutnya slack parameter. Jui Yang Hsia dan Chih-Jen Lin dalam tulisannya[[15]](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/warm-start/svr-param.pdf) menyatakan bahwa regularization parameter ini berfungsi untuk menghindari overfitting pada training data.

Kembali pada kasus prediksi rumah di Boston, mari kita coba tambahkan parameter C pada data. Jika kita set nilai C = 2.0, lantas hasilnya adalah sebagai berikut.

[](https://www.dicoding.com/academies/184/tutorials/8457?from=8452)

Perhatikan bahwa sekarang model kita menyesuaikan sebaran data dengan lebih baik dibanding model sebelumnya.

Sampai sini sudah paham ya? Ada dua parameter yang kita gunakan dalam model SVR, yaitu parameter ϵ yang menunjukkan margin of error dan parameter C yang merupakan parameter keteraturan atau regularization parameter.

Ada tiga jenis implementasi Support Vector Regression pada scikit-learn yaitu: SVR, NuSVR, dan LinearSVR. Implementasi dari LinearSVR lebih cepat dari SVR tetapi hanya dapat digunakan untuk kernel linear, sedangkan NuSVR mengimplementasikan formula yang sedikit berbeda dari SVR dan LinearSVR. NuSVR menggunakan parameter nu untuk mengontrol jumlah support vector. Jika tertarik untuk lebih memahami bagaimana implementasi SVR pada scikit-learn, Anda bisa mengunjungi tautan berikut: [SVR](https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html#sklearn.svm.SVR), [NuSVR](https://scikit-learn.org/stable/modules/generated/sklearn.svm.NuSVR.html#sklearn.svm.NuSVR), dan [LinearSVR](https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVR.html#sklearn.svm.LinearSVR).