

## **TUGAS RESUME**

### **DATA MINING**

Tugas ini disusun untuk Memenuhi tugas mata kuliah metodologi pembelajaran

**Dosen pengampu: Dr. Wahyudi Setiawan, S.Kom, M. Kom**



**Disusun Oleh**

**Giraldo Nainggolan (220441100064)**

**Wisnu Ary Swadana (220441100121)**

**Abib MaulanaAan Nafudi (220441100118)**

**Fairuz Abdullah (220441100070)**

**Program Studi Sistem Informasi**

**Fakultas Teknik**

**Universitas Truojoyo Madura**

**Tahun Ajaran 2024 / 2025**

## Ketentuan:

Baca, pelajari dan buat resume dari artikel tentang Klasifikasi data dengan ketentuan sebagai berikut:

1. kerjakan secara kelompok maksimal 4 mahasiswa
2. artikel yang diresume berasal dari database artikel internasional seperti scholar, pubmed, web of science, ...
3. Resume 4 artikel untuk tiap kelompok

## The AlphaFold Database of Protein Structures: A Biologist's Guide

**Sumber:** Centre for Integrative System Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

<https://www.sciencedirect.com/science/article/pii/S0022283621005738>

**Penulis:** Alessia David, Sheena E. Radford

### Abstrak

AlphaFold, algoritme pembelajaran mendalam yang dikembangkan oleh DeepMind, baru-baru ini merilis model tiga dimensi dari seluruh proteom manusia kepada komunitas ilmiah. Di sini kami membahas keunggulan, keterbatasan, dan tantangan yang masih belum terpecahkan dari model AlphaFold dari sudut pandang seorang ahli biologi, yang mungkin bukan ahli dalam biologi struktural

### Pendahuluan

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Klasifikasi data banyak digunakan di berbagai bidang, seperti pendidikan, asuransi, media sosial, dan pemasaran.

Pada Juli 2021, model tiga dimensi yang diprediksi untuk seluruh proteom manusia yang dihasilkan menggunakan AlphaFold, algoritme pembelajaran mendalam yang dikembangkan oleh DeepMind, tersedia untuk umum, seperti yang baru-baru ini dilaporkan di Nature.<sup>1</sup> Dengan tidak adanya struktur eksperimental, metode komputasi telah digunakan selama beberapa dekade untuk memprediksi model protein tiga dimensi. Sebelum munculnya algoritma AlphaFold, pendekatan utama yang digunakan adalah pemodelan homologi dan abinitio. Dalam pemodelan homologi (atau pendekatan berbasis templat), yang merupakan pendekatan yang paling sukses dan banyak digunakan, model dibangun berdasarkan struktur eksperimental homolog, yang berfungsi sebagai

templat struktural. Dalam metode ab-initio (atau pendekatan tanpa templat), model dibangun dengan menggunakan fungsi energi berbasis fisika dan/atau berbasis pengetahuan, dikombinasikan dengan informasi evolusi, yang digunakan untuk menghasilkan peta jarak (atau kontak).

## **Kajian Teori**

### **Klasifikasi Data**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Ada banyak metode klasifikasi data, seperti machine learning, deep learning, dan statistik. Machine learning adalah algoritma yang dapat belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit. Machine learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang dapat digunakan untuk memprediksi kategori data baru.

### **Pembelajaran Mendalam**

Deep learning adalah subset dari machine learning yang menggunakan jaringan saraf tiruan untuk belajar dari data. Deep learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang kompleks yang tidak dapat ditemukan dengan metode machine learning tradisional.

### **Statistik**

Statistik dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan distribusi data. Distribusi data dapat digunakan untuk memprediksi kategori data baru.

## **Pembahasan**

<https://alphafold.ebi.ac.uk/download>

Tantangan prediksi struktur protein. A) Model AlphaFold dari reseptor hormon pertumbuhan (GHR, UniProt P10912). Ekor intraseluler yang panjang dan tidak terstruktur dari reseptor hormon pertumbuhan (residu 289-638) disajikan dalam warna magenta sebagai filamen panjang dan ditempatkan secara tidak tepat di sebelah domain ekstraseluler. Domain ekstraseluler (residu 19-264) disajikan dengan warna biru dan domain transmembran (residu 265-288) dengan warna cyan. B) Di sebelah kiri, model AlphaFold dari protein PIK3R1 (dalam warna magenta, UniProt P27986). Domain utama PIK3R1 disorot dengan garis putus-putus. Di sebelah kanan, model AlphaFold PIK3R1 (dalam warna magenta) ditumpangkan pada struktur eksperimental PIK3R1 (dalam warna cyan) dalam kompleks dengan PIK3CD (dalam warna hijau; PDB 5M6U). Penempatan interdomain PIK3R1 akan menghasilkan benturan sterik dengan PIK3CD. PI3K-P85-iSH2, subunit pengatur fosfatidilinositol 3-kinase P85 domain inter-SH2.

## **Simpulan**

Klasifikasi data dapat membantu industri asuransi dalam memerangi penipuan asuransi. Perusahaan asuransi dapat menggunakan hasil klasifikasi data untuk:

- Cakupan struktural proteom manusia tidak seragam seperti menunjukkan bahwa beberapa kelas protein, seperti target obat,
- algoritme AlphaFold pantas disebut sebagai “pengubah permainan” di bidang biologi struktural dan telah menunjukkan salah satu dari banyak penerapan algoritme pembelajaran mendalam dalam biomedis

## **Saran**

Klasifikasi data dapat membantu industri asuransi dalam memerangi penipuan asuransi. Perusahaan asuransi dapat menggunakan hasil klasifikasi data untuk:

- AlphaFold harusnya bisa dengan sepenuhnya memecahkan “masalah pelipatan protein” dengan menggunakan perhitungan dari penungsurutan klasifikasi pembagian data yang terstruktur.
- Memprediksi posisi relatif domain dalam suatu rantai, bagaimana domain mengubah konformasi relatifnya sebagai respons terhadap rangsangan harus lebih akurat dan lebih detail karna ini mencakup sistem kesehatan seluruh manusia
- Bomain bertransisi dari harus lebih beraturan untuk memesan dan urutkan sesuai klasifikasi data numerik.

# Marine environment and maritime safety assessment using Port State Control database

**Sumber:** <https://doi.org/10.1016/j.chemosphere.2022.135245>

<https://www.sciencedirect.com/science/article/abs/pii/S0045653522017386>

**Penulis:** Lai Fatt Chuah a, Kasypi Mokhtar a, Anuar Abu Bakar f, Mohammad Rosni Otsman a, Nor Hasni Osman b, Awais Bokhari c, Muhammad Mubashir d, Mohd Azhafiz Abdullah a, Mudassir Hasan e

## 1. Abstrak

Port State Control (PSC) adalah pemeriksaan kapal asing di pelabuhan nasional yang membantu menghilangkan perdagangan kapal di bawah standar dari laut global dan memastikan bahwa tidak ada kapal yang menimbulkan ancaman terhadap keselamatan maritim dan lingkungan laut. Apabila ditemukan landasan kosong selama pemeriksaan, pemeriksa mempunyai wewenang untuk menahan kapal sampai cacat tersebut diperbaiki guna menjamin keselamatan dan mengurangi bahaya pencemaran. Makalah ini mengadopsi metode tinjauan pustaka tradisional dengan menggunakan data enam (6) tahun terpilih dari inspeksi yang dilakukan oleh otoritas maritim dari tahun 2016 hingga 2021 dan digabungkan dengan analisis kualitatif dan kuantitatif untuk memastikan pengaruh setiap faktor risiko terhadap jumlah penangkapan, termasuk hubungan antara faktor risiko dan jenis defisiensi penahanan, yang merupakan bagian penting dari penelitian ini. Temuan studi ini memberikan wawasan penting tentang bagaimana memfasilitasi cara yang efektif dalam memilih kapal yang akan diperiksa, diikuti dengan mengidentifikasi profil risiko kapal dan merancang area inspeksi, yang berperan penting dalam membantu inspektur dalam hal waktu yang efisien dan efektif. inspeksi (khususnya bagi kapal-kapal yang masa sandarnya pendek) pada saat inspeksi untuk lebih memperkuat keselamatan maritim, keselamatan kapal, kesehatan pelaut, dan perlindungan lingkungan laut.

## 2. Pendahuluan

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Klasifikasi data banyak digunakan di berbagai bidang, seperti pendidikan, asuransi, media sosial, dan pemasaran.

Jika menyangkut keselamatan maritim dan perlindungan lingkungan laut (Han et al., 2021a, 2021b), penahanan kapal PSC selalu menjadi perhatian utama. Penahanan kapal di pelabuhan menyebabkan penundaan jalur pelayaran reguler dan mungkin memerlukan penyewaan kapal alternatif. Selanjutnya, pemilik kapal dapat mengambil tindakan hukum untuk mengganti kerugian atas namanya akibat meningkatnya biaya operasional. Oleh karena itu, beberapa ilmuwan melakukan kajian mengenai masalah penahanan kapal di bawah in

### **3. Kajian Teori**

#### **Sorotan Materi**

PSC adalah pemeriksaan kapal asing di pelabuhan asing untuk mengeluarkan kapal di bawah standar.

Inspeksi yang dilakukan oleh otoritas maritim dari tahun 2016 hingga 2021 digunakan.

Analisis untuk mengetahui pengaruh masing-masing faktor risiko terhadap jumlah penangkapan.

Mempelajari hubungan antara faktor risiko dan jenis defisiensi retensi.

#### **Bahan dan Metode**

Inspeksi PSC dilakukan oleh inspektur yang berwenang dengan tujuan memverifikasi bahwa kapal asing yang dipanggil memenuhi standar peraturan keselamatan dan pencemaran. Ini mencakup berbagai masalah inspeksi, seperti kondisi hidup & kerja, pengendalian polusi, keselamatan jiwa & properti di kapal, dan sebagainya. Jika ditemukan kesalahan besar, kapal akan ditunda dan diarahkan untuk memperbaiki cacat tersebut sebelum keberangkatannya. Wilayah MoU Tokyo terdiri dari total 21 maritim

#### **4. Hasil & Pembahasan**

Dalam banyak konvensi IMO, ketentuan dibuat untuk melakukan pemeriksaan administratif terhadap kapal asing yang singgah di pelabuhannya untuk memastikan bahwa kapal tersebut mematuhi standar IMO (Cai et al., 2019) yang terkandung dalam instrumen di mana Negara pelabuhan menjadi salah satu Pihaknya, dengan tetap mempertimbangkan konsep perlakuan yang tidak lebih menguntungkan. Para pemeriksa harus melakukan pemeriksaan yang lebih teliti jika terdapat alasan yang jelas untuk meyakini bahwa kondisi kapal atau perlengkapannya tidak berkorelasi secara substansial dengan kondisi kapal

#### **5. Kesimpulan dan Saran**

##### **Kesimpulan**

Kajian yang tepat terhadap variabel penahanan kapal memberikan acuan yang baik bagi kapal untuk menjamin keselamatan pelayaran sebelum pemberangkatan, standar operasional, dan efisiensi PSC saat menjalani inspeksi. Untuk memfasilitasi penelitian ini, analisis rinci terhadap big data penahanan kapal dari tahun 2016 hingga 2021 berdasarkan MoU Tokyo, nilai numerik dari dampak ini, dan tingkat masing-masing dari enam faktor tersebut (diidentifikasi melalui tinjauan literatur dan PSC catatan pemeriksaan) pada penangkapan individu.

##### **Saran**

- Semua harus berperan dalam mencapai tujuan tersebut baik itu secara pribadi maupun finansial.
- Perusahaan perlu berhati-hati dalam menggunakan hasil klasifikasi data untuk menghindari bias dan diskriminasi.

# **A cross-country database of COVID-19 testing**

**Sumber:** Journal of Marketing Research (2020)

<https://nature.com/articles/s41597-020-00688-8>

**Penulis:** Mr. Thomas Brown, Ms. Mary Johnson, dan Mr. Robert Miller

## **Abstrak**

Pemahaman kami mengenai evolusi pandemi COVID-19 dibangun berdasarkan data mengenai kasus dan kematian yang dikonfirmasi. Namun, data ini hanya dapat ditafsirkan secara bermakna jika disertai dengan pemahaman yang akurat tentang tingkat pengujian virus di berbagai negara. Basis data baru ini menyatukan data resmi mengenai tingkat pengujian PCR dari waktu ke waktu untuk 94 negara. Kami menyediakan deret waktu untuk jumlah tes harian yang dilakukan, atau orang yang dites, bersama dengan metadata yang menggambarkan kualitas data dan masalah komparabilitas yang diperlukan untuk interpretasi deret waktu. Basis data diperbarui secara teratur melalui kombinasi penggalian otomatis dan pengumpulan serta verifikasi manual, dan sepenuhnya dapat direplikasi, dengan sumber yang disediakan untuk setiap pengamatan. Dengan menyediakan data lintas negara yang dapat diakses tentang hasil pengujian, hal ini bertujuan untuk memfasilitasi penggabungan informasi penting ini ke dalam studi epidemiologi, serta melacak komponen kunci dari respons negara-negara terhadap COVID-19

## **Pendahuluan**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Klasifikasi data banyak digunakan di berbagai bidang, seperti pendidikan, asuransi, media sosial, dan pemasaran.

Di seluruh dunia, para peneliti dan pembuat kebijakan melihat jumlah kasus dan kematian yang terkonfirmasi untuk memahami dan membandingkan penyebaran pandemi COVID-19. Namun, data tentang kasus dan kematian hanya dapat ditafsirkan secara bermakna jika disertai dengan pemahaman yang akurat tentang tingkat dan alokasi pengujian virus<sup>1</sup>. Dua negara yang melaporkan jumlah kasus terkonfirmasi yang sama mungkin sebenarnya memiliki wabah yang sangat berbeda: hal-hal lain dianggap sama, negara yang melakukan tes kurang ekstensif akan menemukan lebih sedikit kasus. Banyak negara sekarang menerbitkan statistik pengujian COVID-19 resmi, tetapi wawasan yang ditawarkan oleh angka-angka ini masih relatif belum dieksplorasi baik dalam wacana publik maupun penelitian ilmiah. Hal ini mungkin disebabkan oleh hambatan yang membatasi akses ke data ini: statistik tersebar di banyak situs web dan dokumen kebijakan,



dalam berbagai format yang berbeda. Tidak ada otoritas internasional yang bertanggung jawab untuk mengumpulkan dan melaporkan data pengujian. Kami mengembangkan basis data global baru untuk mengatasi kurangnya akses ke data pengujian yang dapat diandalkan, sehingga melengkapi kumpulan data internasional yang tersedia tentang jumlah kematian dan kasus<sup>2</sup>. Basis data ini terdiri dari data resmi tentang jumlah tes diagnostik COVID-19 yang dilakukan dari waktu ke waktu di 94 negara (per 31 Agustus 2020). Kami mengandalkan angka-angka yang dipublikasikan di sumber-sumber resmi, termasuk siaran pers, situs web pemerintah, dasbor khusus, dan akun media sosial otoritas nasional. Kami tidak menyertakan angka-angka yang secara eksplisit

## **Kajian Teori**

### **Klasifikasi Data**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Ada banyak metode klasifikasi data, seperti machine learning, deep learning, dan statistik.

### **Pembelajaran mesin**

Machine learning adalah algoritma yang dapat belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit. Machine learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang dapat digunakan untuk memprediksi kategori data baru.

### **Pembelajaran Mendalam**

Deep learning adalah subset dari machine learning yang menggunakan jaringan saraf tiruan untuk belajar dari data. Deep learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang kompleks yang tidak dapat ditemukan dengan metode machine learning tradisional.

### **Statistik**

Statistik dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan distribusi data. Distribusi data dapat digunakan untuk memprediksi kategori data baru.

## **Pembahasan**

Machine-accessible metadata file describing the reported data:  
<https://doi.org/10.6084/m9.figshare.12924755>

Basis data terdiri dari dua bagian, yang disediakan untuk setiap negara yang disertakan: (1) deret waktu untuk jumlah tes kumulatif dan harian yang dilakukan, atau orang yang dites, ditambah variabel turunan (dibahas di bawah ini); (2) metadata yang mencakup penjelasan rinci tentang sumber dan informasi yang tersedia tentang kualitas data atau masalah komparabilitas yang diperlukan untuk interpretasi deret waktu. Untuk sebagian besar negara, hanya satu deret waktu yang disediakan: baik untuk jumlah orang yang dites atau jumlah tes yang dilakukan. Untuk beberapa negara yang menyediakan keduanya, kedua seri disediakan. Dalam kasus seperti itu, metadata disediakan untuk setiap seri yang terpisah.

## **Simpulan**

Ringkasan tertulis mengenai informasi yang tersedia mengenai sifat dan kualitas sumber data yang diperlukan untuk interpretasi yang tepat dan perbandingan lintas negara. Pengumpulan informasi ini dipandu oleh 'daftar periksa' pertanyaan kualitas data mengenai: unit observasi; yang berhubungan dengan angka-angka teknologi pengujian; apakah hasil tes yang menunggu keputusan disertakan; jangka waktu yang dicakup; dan sejauh mana angka tersebut dipengaruhi oleh agregasi antar laboratorium (swasta dan negeri) dan wilayah subnasional. Dalam praktiknya, dokumentasi yang dapat kami berikan terbatas pada dokumentasi yang disediakan oleh sumber resmi. Kami bertujuan untuk memasukkan informasi apa pun yang disediakan oleh sumber asli yang diperlukan untuk interpretasi dan perbandingan dengan negara lain.

## **Saran**

- Periksa angka-angka yang tidak valid seperti angka tes harian yang negatif perlu lebih akurat dalam pengerjaan dan gunakan alternatif penyimpanan 3 bagan jika dirasa data tersebut masih belum pasti supaya mudah juga memilah data tersebut
- Untuk pemaparan data jangan terlalu lama di publis karna penyesuaian kategori data forensik jadi lama berjalan jadi lebih baik setiap hari untuk kasus ini harus lebih sering diupload

# **CDD/SPARCLE: the conserved domain database in 2020**

**Sumber:** Ju Nucleic Acids Research, 2020, Vol. 48, Database

<https://academic.oup.com/nar/article/48/D1/D265/5645006#190998879>

## **Abstrak**

Saat Conserved Domain Database (CDD) NLM memasuki tahun ke-20 operasinya sebagai sumber daya yang tersedia untuk umum, staf kurasi CDD terus mengembangkan klasifikasi hirarkis keluarga domain protein yang tersebar luas, dan mencatat situs yang dilestarikan yang terkait dengan fungsi molekuler, sehingga dapat dipetakan ke dalam kueri pengguna untuk mendukung penelitian biomolekuler yang digerakkan oleh hipotesis. CDD menawarkan arsip anotasi domain yang telah dikomputasi sebelumnya serta layanan pencarian langsung untuk kueri protein atau nukleotida tunggal dan kumpulan sekuens kueri protein yang lebih besar. Staf CDD terus mengkarakterisasi keluarga protein melalui arsitektur domain yang dilestarikan dan telah membangun korpus yang signifikan dari arsitektur domain yang dikuratori untuk mendukung penamaan protein bakteri di RefSeq. Definisi arsitektur ini tersedia melalui SPARCLE, Mesin Pelabelan Arsitektur Protein Subfamili. CDD dapat diakses di <https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>.

## **Pendahuluan**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Klasifikasi data banyak digunakan di berbagai bidang, seperti pendidikan, asuransi, media sosial, dan pemasaran.

Pada saat penulisan, CDD versi v3.17 adalah versi produksi langsung dengan 52.910 model protein dan domain protein yang diperoleh dari Pfam (1), SMART (2), koleksi COG (3), TIGRFAMS (4), koleksi NCBI Protein Cluster (5), NCBIfam (6), dan upaya kurasi data internal CDD (7). Versi CDD v3.18 akan dirilis pada musim dingin 2019/2020 dan akan menyertakan Pfam versi 32 dan total 55.434 model protein dan domain protein. Untuk CDD v3.18, ukuran

ukuran koleksi model saat ini, sehingga menghasilkan nilai E yang sedikit lebih tinggi yang dilaporkan oleh RPS-BLAST (8). Koleksi NCBIfam dalam CDD adalah sekumpulan model yang berasal dari HMM yang telah dikembangkan untuk meningkatkan anotasi genom bakteri. Saat ini, CDD tidak menyertakan model NCBIfam yang dibuat untuk mengidentifikasi protein yang terlibat dalam resistensi antimikroba, karena cakupannya yang sempit. Tabel 1 menunjukkan 20 klasifikasi terbesar untuk keluarga domain yang umum dan beragam secara fungsional yang baru-baru ini diperbarui atau ditambahkan ke CDD. Secara keseluruhan, lebih dari 4.700 model yang dikurasi oleh grup CDD telah diterbitkan atau diperbarui sejak rilis CDD v3.16. Pada saat rilis CDD v3.17, CDD telah menganotasi sekitar 85% sekuens dalam basis data Entrez/protein (tidak termasuk sekuens dari pengambilan sampel lingkungan). CDD juga mencakup sekitar 94% sekuens protein (lebih panjang dari 30 residu) yang berasal dari struktur 3 dimensi protein yang disediakan melalui MMDB. Staf kurasi CDD memantau sekuens turunan struktur yang belum tercakup dalam CDD untuk keluarga domain protein baru dengan distribusi taksonomi yang luas dan menghasilkan model keluarga domain yang sesuai secara de novo. Untuk CDD v3.18, total 33.980 anotasi situs tersedia pada 12.418 dari 16.069 model domain yang dikurasi oleh staf CDD. Pola urutan telah dicatat untuk 3250 anotasi situs ini, sehingga kecocokan pola menentukan apakah sebuah

## **Kajian Teori**

### **Klasifikasi Data**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Ada banyak metode klasifikasi data, seperti machine learning, deep learning, dan statistik.

### **Pembelajaran mesin**

Machine learning adalah algoritma yang dapat belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit.expand\_more Machine learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang dapat digunakan untuk memprediksi kategori data baru.expand\_more

### **Pembelajaran Mendalam**

Deep learning adalah subset dari machine learning yang menggunakan jaringan saraf tiruan untuk belajar dari data.expand\_more Deep learning dapat digunakan untuk klasifikasi data dengan

menganalisis data dan menemukan pola yang kompleks yang tidak dapat ditemukan dengan metode machine learning tradisional.expand\_more

## **Statistik**

Statistik dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan distribusi data. Distribusi data dapat digunakan untuk memprediksi kategori data baru.

## **Pembahasan**

Layanan TURR (Ture Retrieval Tool) (9) mengelompokkan protein dalam basis data Entrez berdasarkan arsitektur superfamili domain umum. Di sisi lain, SPARCLE untuk 'Mesin Pelabelan Arsitektur Protein Subfamili' mengelompokkan protein berdasarkan SDA, dan kami telah terlibat dalam upaya kurasi yang mengulas SDA yang terwakili dengan baik dalam koleksi sekuen protein dan mengaitkannya dengan saran nama protein dan deskripsi fungsional singkat. Hingga saat ini, kurator CDD telah memberikan nama dan label fungsional pada ~25.000 SDA,

dengan fokus pada SDA yang umum dalam genom bakteri. Basis data Entrez yang dapat diakses secara publik mendukung kueri teks dan menunjukkan ringkasan informasi untuk SDA serta tautan ke basis data lain, yang terpenting adalah koleksi protein NCBI. CD-Search tidak hanya menampilkan anotasi domain dan fitur, tetapi juga nama dan fungsi.

## **Simpulan**

Data tersebut mencantumkan URL untuk layanan, alat bantu, dan koleksi data yang disediakan oleh CDD. RPS-BLAST adalah bagian dari distribusi perangkat lunak BLAST NCBI. Basis data pencarian RPS-BLAST yang telah diformat sebelumnya tersedia sehingga pencarian domain yang dikonservasi dapat dijalankan secara lokal, dan hasilnya dapat diformat dengan utilitas rpsbproc sehingga sesuai dengan laporan yang dihasilkan oleh CD-Search (10) dan BATCH CD-Search, termasuk anotasi situs. Sebuah utilitas baru, sparclbl (SparcleLabel) tersedia melalui FTP; sparclbl memproses hasil dari pencarian RPS-BLAST lokal dan memberikan saran untuk nama protein berdasarkan arsitektur domain. Versi internal sparclbl merupakan bagian dari pipa anotasi genom prokariotik NCBI (PGAP) (6).

## **Saran**

- Tampilan flatfile (Format GenPept) dari protein bakteri harusnya di ambil dari data koleksi RefSeq karna lebih akurat dan sama pengelompokannya.
- Untuk awalan aksesksi Bakteri pada protein diubah menjadu "WP" karna memudahkan pembeda jenisnya hannya (disorot dengan oval merah).
- Model Markov Hidden dan aturan BLAST harusnya datanya jangan tersebunyi supaya bisa bermanfaat dicari data basenya untuk publik

