Demo Day Proposal

**Machine Learning with TensorFlow Training**

**Professional Academy Digital Talent Scholarship 2022**

|  |  |
| --- | --- |
| **Group Number** | **HC\_1** |
| **Dataset** | **Healthcare** |
| **Name - DTS ID** | 1. **Anthonius Adi Nugroho (152236035100-32)** 2. **Alfonsus Haryo Sangaji (152236035101-664)** 3. **Alma Vita Sophia (152236035101-224)** 4. **Alda Raharja (152236035101-478)** 5. **Zia Firnanda (152236035100-942)** |

**Selected Theme**: Health Care

**Title of the Project**: Disease Classification Using Neural Network

**Latar belakang masalah**

Perkembangan teknologi AI sangat pesat khususnya dalam bidang data. Data dipandang mulai berguna dalam membantu menentukan pilihan. Beberapa tahun terakhir perkembangan teknologi mempercepat proses terjadinya inovasi di berbagai lini, termasuk inovasi teknologi di bidang kesehatan. Salah satu platform memiliki banyak peminat perkembangan teknologi kesehatan masa depan adalah kecerdasan buatan atau artificial intelligence (AI). Perkembangan dalam dunia medis ini dapat membantu manusia untuk mendapatkan insight baru yang yang membantu pekerja medis.

Di bidang kedokteran, kecerdasan buatan dapat diaplikasikan dalam hal diagnosis, tatalaksana, dan prediksi risiko. Kecerdasan buatan diharapkan dapat melakukan hal-hal berikut:

 Membantu klinisi mendiagnosis suatu penyakit dan mengoptimalkan proses tatalaksana

 Mengurangi angka misdiagnosis dan meningkatkan efisiensi diagnostik

 Mengenali hasil pencitraan dan memberi informasi diagnostik pencitraan yang lebih akurat

 Menyediakan hasil analisis prediksi pasien yang lebih akurat menggunakan analisis big data

 Mendukung penelitian obat-obatan dan meningkatkan efisiensi pengembangan obat baru

Pada penelitian ini mencoba melakukan klasifikasi 41 penyakit dari 132 gejala yang yang dialami oleh pasien. Data gejala penyakit memiliki bobot sudah disediakan yang akan menjadi acuan dalam menentukan penyakit yang dialami oleh pasien. Dalam menentukan penyakit yang dialami oleh pasien, dilakukan pelatihan menggunakan data yang sudah ada dengan neural network. Hasil dari pelatihan akan menghasilkan model yang dapat mengklasifikasikan 41 penyakit dengan gejala yang di inputkan.

**Pernyataan Masalah**

Menentukan penyakit berdasarkan gejala yang dialami dan memberikan saran pengobatan yang diberikan

**Pertanyaan Penelitian**

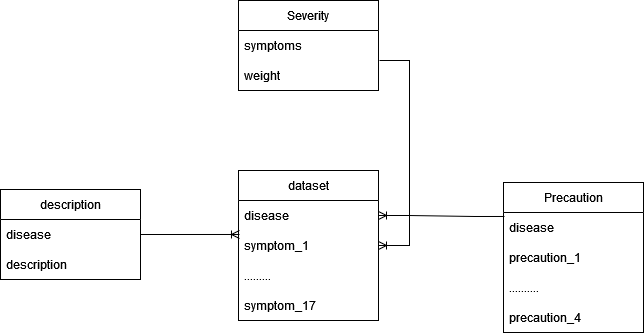
Bagaimana Menentukan 41 penyakit berdasarkan gejala yang yang dialami dan memberikan saran pengobatan yang diberikan

**Project Scope & Deliverables**

Masing masing anggota membuat data cleaning dan preprocessing, hasil data yang sudah clean dari masing masing anggota akan didiskusikan dan akan disepakati yang akan digunakan untuk tahap selanjutnya. Tahap selanjutnya akan dilakukan pelatihan neural network menggunakan tensorflow. Hasil akhir dari penelitian ini dapat melakukan klasifikasi penyakit dari gejala yang di inputkan oleh user. Performa model akan diukur saat melakukan testing. Penjelasan diatas dapat digambarkan dalam flowchart seperti dibawah ini.



Dataset yang dimiliki memiliki 4 tabel, yaitu tabel dataset, precaution, description dan severity. Tabel dataset berisi record dari pasien, tabel precaution berisi pencegahan dari disease yang dialami pasien, tabel description berisi deskripsi dari penyakit/disease dan tabel severity merupakan tabel yang berisi bobot dari symptom/gejala yang dialami. Relasi antar tabel dapat dibuat kardinalitasnya seperti gambar dibawah ini.



**Project Schedule:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|  |  |  |  |  |  |  |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|  |  |  |  |  |  |  |
| 24 | 25 | 26 | 27 |  |  |  |
|  |  |  |  |  |  |  |

**Keterangan**

kuning : Zoom Meeting

Hijau : Pengumpulan

Cokelat : Pengerjaan

**Schedule:**

10: zoom pertama (pengenalan dan diskusi dataset)

11-13: masing anggota melakukan data cleaning dan eksplorasi data

14: mendiskusikan hasil pekerjaan dan rancangan proposal

15: pengerjaan proposal

16: mendiskusikan proposal

17-18: pembuatan model

19: mengevaluasi model

20-22: melakukan pembaharuan model dan deploy model

23: mengevaluasi dan mendiskusikan keseluruhan model dan rancangan laporan akhir

24-25: melakukan penyempurnaan model dan pembuatan laporan akhir serta video

26: evaluasi dan pertemuan terakhir

**Risk and Issue Management Plan:**

* Fitur yang tidak ada bobotnya
* Fitur yang memiliki bobot ganda atau redundant
* Banyaknya data duplikasi pada dataset (kasus setiap penyakit dalam dataset sedikit). Data pada dataset.csv terdapat **4920**, dan jika data duplikasi dihapus jumlahnya menjadi **304**. Sehingga setiap penyakit memiliki jumlah data/kasus yang sedikit (data kasus penyakit tertinggi hanya 10, dan terendah 5)

**Project Reference:**

<https://www.yarsi.ac.id/2022/04/24/melihat-peran-artificial-intelligence-dalam-dunia-kesehatan/>

<https://www.kaggle.com/datasets/itachi9604/disease-symptom-description-dataset>

<https://www.alomedika.com/peran-artificial-intelligence-dalam-kedokteran-kardiovaskular>