



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kecerdasan buatan merupakan suatu teknologi yang sedang banyak diperbincangkan penggunanya. Teknologi ini sudah banyak membantu masyarakat luas dalam beberapa tahun ini. Produk yang menggunakan kecerdasan buatan juga sudah mulai bermunculan baik di bidang kesehatan, interaksi dengan manusia, industri, dan bidang lainnya (Anthony, n.d.). Kecerdasan buatan memiliki berbagai macam cabang, salah satunya adalah pemelajaran mesin (*Machine Learning*). Pemelajaran mesin adalah suatu sistem yang dapat secara otomatis mengeluarkan hasil dari pemelajaran dan belajar memperbaiki diri dari pengalaman tanpa perlu diprogram ulang (Sodhi et al., 2019). Untuk dapat menghasilkan suatu keluaran yang diinginkan, pembelajaran mesin membutuhkan data masukan. Data yang digunakan untuk pemrosesan pemelajaran mesin tidak sedikit dan beragam mulai dari gambar, teks, suara, hingga video.

Salah satu topik yang menarik adalah *speech recognition*, pada tahun 2019 salah satu website penjualan online yaitu amazon mendapatkan keuntungan sebesar \$3.03 miliar pada kuartal yang sama tahun sebelumnya. Penjualan terbanyak diraih oleh *echo device* seperti alexa, siri, dan *device* lainnya yang menjadikan mereka sebagai salah satu perangkat suara terlaris hingga saat ini (Helmore, 2019). Google mengklaim bahwa 20% penelusuran sudah dilakukan melalui suara. Diperkirakan 30% pencarian pada tahun 2020 akan dilakukan tanpa layar (google assistant, cortana, dll) (Greening, 2018). Setidaknya perkembangan pengenalan suara pada tahun 2024 mencapai \$21.5 miliar dengan CAGR 19.18% (Market, 2019).

Dengan perkembangan speech recognition yang semakin maju, penelitian tentang speech recognition semakin banyak dikaji di berbagai bahasa. Namun, sangat sedikit dataset suara *open source* dalam bahasa jepang di berbagai website. Dataset suara dalam bahasa jepang pada website appen.com sejumlah 33 jam dengan 144 suara. Per tanggal 30 September 2020 setidaknya hanya 7 jam suara yang valid dari 9 jam suara pada website common voice

(<https://commonvoice.mozilla.org/>). Hal ini membuat sulitnya penelitian sehingga diperlukan suatu cara untuk mendapatkan dataset.

Salah satu cara yang dapat digunakan untuk mengumpulkan dataset tersebut adalah dengan pembuatan website. Banyak pilihan bahasa pemrograman untuk membuat suatu website salah satunya adalah php. Php merupakan bahasa pemrograman yang banyak dipakai dalam pembuatan website, karena keunggulannya yang mudah dipahami (Kamaruzzaman, 2020). Selain itu php juga dapat menggunakan *framework* sehingga memudahkan programmer untuk membuat suatu website. Sudah cukup banyak *framework* php di dunia ini, mulai dari yang paling mudah dipelajari hingga yang lebih modern. Salah satu framework yang menarik adalah laravel. Laravel memiliki beberapa keunggulan yaitu kode pemrograman lebih singkat, mudah dimengerti, dan lebih ekspresif.

Universitas Multimedia Nusantara sebagai salah satu perguruan tinggi di Indonesia yang berfokus pada teknologi informasi dan komunikasi. Membina mahasiswa untuk menjadi pribadi yang berwawasan internasional dan berkemampuan tinggi serta berjiwa wirausaha. UMN juga memberikan kesempatan kepada mahasiswa serta tenaga pengajar untuk mengembangkan teknologi informasi agar dapat membantu masyarakat luas. Salah satu cara untuk mengembangkan teknologi informasi adalah penelitian tentang topik tertentu, salah satu dosen Informatika, Bapak Julio Christian Young melakukan penelitian untuk *speech recognition* dalam bahasa jepang. Untuk merealisasikan penelitian itu, dibutuhkan dataset dalam bentuk suara dan untuk mempermudah pengumpulan dataset diperlukan sebuah website pengumpulan data suara.

Dengan topik yang menarik dan manfaat yang besar, menjadi pilihan yang bagus untuk praktik kerja magang. Oleh sebab itu dibuatlah sebuah website platform pengumpulan speech dataset bahasa jepang dengan menggunakan bahasa pemrograman php dan *framework* laravel sehingga diharapkan pengumpulan dataset lebih mudah dan tidak memakan banyak waktu. Diharapkan dengan adanya website ini dapat membantu mengumpulkan dataset yang nantinya akan digunakan untuk penelitian dan dengan adanya website ini diharapkan dapat membantu banyak orang.

1.2 Maksud dan Tujuan Kerja Magang

Adapun maksud dan tujuan dari kerja magang ini yaitu untuk beradaptasi dengan dunia kerja, menambah pengalaman, mengaplikasikan ilmu yang telah dipelajari selama perkuliahan, menganalisis kegiatan serta persoalan yang dihadapi dan berhubungan dengan bidang keilmuan, dan sebagai sarana pelatihan kerja bagi mahasiswa untuk memperoleh bekal di dunia kerja. Disamping itu, kegiatan magang ini bertujuan untuk membuat Website platform pengumpulan speech dataset untuk bahasa jepang dengan menggunakan php.

1.3 Waktu dan Prosedur Pelaksanaan Kerja Magang

Waktu pelaksanaan kerja magang ini dilaksanakan selama 90 hari kerja terhitung dari tanggal 03 Agustus 2020 sampai dengan 03 November 2020 dengan pembimbing lapangan oleh Julio Christian Young, selaku PIC Lab. Artificial Intelligence Universitas Multimedia Nusantara, Gading Serpong, Tangerang. Adapun prosedur pelaksanaan kerja magang di Universitas Multimedia Nusantara adalah sebagai berikut.

- a. Mahasiswa melamar magang melalui *email*, kemudian menunggu hasil pengumuman lebih lanjut.
- b. Mahasiswa melakukan kegiatan magang secara *daring* selama 5 hari, yaitu pada hari Senin s.d. Jumat.
- c. Lama jam kerja adalah delapan jam per hari. Perhitungan jam dimulai dari jam masuk.
- d. Mahasiswa melakukan interview mengenai *soft skill* sebagai salah satu syarat mengikuti kegiatan magang di Universitas Multimedia Nusantara.