

## PENGEMBANGAN KAMUS ISTILAH BASIS DATA BERBASIS WEBSITE MENGUNAKAN ALGORITMA COSINE SIMILARITY UNTUK MENINGKATKAN HASIL BELAJAR SISWA

**Rahmadina Puspa Nuristiqomah**

S1 Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
E-mail : rahmadina.17050974028@mhs.unesa.ac.id

**Yeni Anistyasari**

S1 Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
E-mail : yenian@unesa.ac.id

### Abstrak

Pengembangan kamus istilah basis data berbasis *website* dikembangkan menggunakan algoritma *cosine similarity* dan menggunakan bahasa *python* dengan *framework Flask*. Kamus digital ini berisi tentang istilah basis data dan pokok materi tipe data pada basis data. Tujuan penelitian ini untuk mengetahui pengembangan kamus berbasis *website* menggunakan algoritma *cosine similarity* dan mengetahui hasil belajar siswa SMK khususnya program kejuruan RPL setelah menggunakan kamus istilah ini. Peneliti menggunakan desain penelitian *One group pre-test post-test design* dengan metode pengembangan aplikasi ADDIE (*Analyze, Design, Development, Implementation, Evaluation*). Dari hasil uji kelayakan ahli, kamus istilah basis data mendapatkan presentase 80,5% dengan kategori layak digunakan dalam proses pembelajaran. Berdasarkan hasil penelitian, respon siswa sangat positif dalam menggunakan media pembelajaran ini. Dapat dilihat dari peningkatan rata – rata nilai post-test dan pre-test, hasil belajar siswa meningkat setelah menggunakan media pembelajaran kamus istilah basis data.

**Kata kunci :** kamus digital, *cosine similarity*, *python*, basis data

### Abstract

*The development of a website-based database term dictionary was developed using cosine similarity algorithm and using the Python language with the Flask framework. A digital dictionary contains database terms and the subject matter of data types in databases. The purpose of this study was to determine the development of a web-based dictionary using the cosine similarity algorithm and to determine the learning outcomes of vocational students, especially the RPL vocational program after using this dictionary. The researcher uses One group pre-test post-test design with the ADDIE application development method (Analyze, Design, Development, Implementation, Evaluation). Based on the research, student responses were very positive in using this learning media. It can be seen from the increase in the average post-test and pre-test scores, student learning outcomes increased after using the database term dictionary.*

**Keywords:** *digital dictionary, cosine similarity, python, database*

### PENDAHULUAN

Inovasi proses pembelajaran di Indonesia berkembang sangat pesat. Inovasi pembelajaran dapat berupa inovasi metode pembelajaran, model pembelajaran atau inovasi media pembelajaran. Perkembangan Teknologi Informasi mendorong *stakeholder* di bidang pendidikan untuk terus berinovasi menciptakan pembelajaran yang interaktif dan efektif. Guru sebagai fasilitator siswa dituntut untuk memberi fasilitas pembelajaran pada siswa supaya mereka mempunyai pengalaman belajar yang nyata. Dengan

adanya fasilitas tersebut, guru mengajak dan membawa seluruh peserta didik untuk berpartisipasi agar pembelajaran menjadi interaktif dan menyenangkan.

Berdasarkan hasil penelitian yang berjudul “*Effects of Digital Learning Materials on nursing students’ mathematics learning, self-efficacy, and task value in vocational education*” oleh Zwart dkk. (2020), pembelajaran yang menggunakan media digital membuat siswa mengalami penurunan kepercayaan diri saat belajar dan memberi dampak pada hasil belajar siswa yang tidak

ada perbedaan setelah menggunakan media digital. Penelitian ini membuktikan adanya peningkatan hasil belajar siswa setelah menggunakan media pembelajaran kamus istilah basis data. Media digital yang digunakan berbasis web pada *platform Wikispaces* yang mempunyai beberapa fitur seperti, konten belajar, tujuan belajar, penugasan, papan diskusi dan sebagainya. Siswa memperoleh konten belajar berupa video dan dapat berinteraksi secara langsung dengan guru maupun siswa lain pada papan diskusi.

Hal tersebut menunjukkan bahwa perlu dilakukan inovasi pada penggunaan media digital khususnya berbasis web pada proses pembelajaran. *Website* yang digunakan pada penelitian tersebut mempunyai fitur yang masih sangat umum dan kurang menarik sehingga membuat siswa tidak percaya diri saat belajar. Selain itu konten yang disajikan perlu diperbaiki agar siswa tidak merasa jenuh dan enggan memperhatikan penyampaian materi pada saat proses pembelajaran.

Salah satu penelitian yang berjudul *“Improving the performance of dictionary-based approaches in protein name recognition”* oleh Tsuruoka & Tsujii (2004), menciptakan sebuah *search engine* berbasis web berupa kamus yang dapat melakukan pencarian dan penyaringan ID berdasarkan nama protein yang dimasukkan ke dalam program. Pada program yang dibuat, penulis menggunakan algoritma *Naïve Bayes* untuk diimplementasikan ke *search engine* yang dibuat. Pada pembuatan program, penulis memaparkan dua tahap implementasi algoritma *Naïve Bayes*. Yang pertama yaitu teknik pencarian *string*, kemudian yang kedua yaitu memperluas pencarian pada *database* kamus. Hasil penelitian menunjukkan filter pencarian pada *website* yang menggunakan *Naïve Bayes* mengalami peningkatan presisi sehingga beberapa kata yang hilang dan tidak muncul pada saat pencarian sangat sedikit. Algoritma *Naïve Bayes* selain dapat digunakan untuk fitur pencarian juga dapat memberi filter pada hasil pencarian.

Berdasarkan hasil studi literatur dan permasalahan yang telah dipaparkan, penulis ingin melakukan penelitian tentang inovasi media digital berupa kamus pada proses pembelajaran dengan menerapkan teknologi *search engine* menggunakan algoritma *cosine similarity*. Kamus yang akan dikembangkan memiliki fitur pencarian yang dapat memudahkan siswa mencari istilah basis data.

Kamus dengan perantara komputer telah menjadi alat bantu yang penting dan banyak digunakan pada pembelajaran *online*. Namun, terlepas dari kemudahan kamus yang dapat diakses melalui komputer dalam mengambil makna kata, pengguna juga dapat mengurangi waktu yang dihabiskan pembaca untuk membaca setiap kata pada buku atau modul. Selain itu, ukuran kosakata

pembaca merupakan faktor kunci yang memengaruhi proses pencarian, dan efektivitasnya (Chang dkk., 2018). Kamus istilah basis data mempermudah siswa melakukan pencarian materi dan modul pada pembelajaran basis data khususnya pada materi tipe data sehingga dapat membantu siswa memahami istilah basis data dengan kata yang dimasukkan ke dalam program kemudian akan keluar istilah dan arti yang sama atau sinonim dari kata yang telah dimasukkan. Dengan adanya inovasi tersebut diharapkan siswa lebih percaya diri dalam proses pembelajaran sehingga siswa akan mengalami peningkatan hasil belajar.

Melalui penelitian yang berjudul **“Pengembangan Kamus Istilah Basis Data berbasis *Website* menggunakan Algoritma *Cosine Similarity* untuk Meningkatkan Hasil Belajar Siswa Kelas XI RPL di SMKN 1 Surabaya”** akan dicari suatu bukti bahwa inovasi *search engine* berbasis *web based learning* diharapkan dapat memberi dampak pada peningkatan hasil belajar siswa dan proses pembelajaran menjadi interaktif dan efektif.

#### **Pengembangan Perangkat Lunak**

Menurut penelitian *International Business Machine* atau IBM, pengembangan perangkat lunak merupakan serangkaian aktivitas ilmu komputer yang meliputi proses pembuatan, perancangan, dan penerapan yang menghasilkan sebuah program atau *software*. Sedangkan pengertian perangkat lunak yaitu sekumpulan instruksi program untuk menyampaikan informasi berupa perintah kepada komputer apa yang harus dilakukan.

#### **Pembelajaran Berbasis *Website***

*Web Based Learning* merupakan kegiatan pembelajaran jarak jauh dimana dalam pelaksanaannya melalui *website* dan jaringan internet dimana media *e-learning* yang digunakan memiliki sistem yang membantu proses pembelajaran dengan lebih memanfaatkan *website* dengan dukungan koneksi internet. (Wena, 2009: 215). Dengan adanya dukungan teknologi komputer, *handphone* dan jaringan internet saat ini, *website* akan sangat mudah diakses dan fleksibel digunakan dimanapun dan kapanpun.

#### **Kamus Digital**

Menurut KBBI (2020), Kamus adalah acuan atau karya rujukan dalam bentuk digital maupun cetak yang meliputi kata dan ungkapan dan dapat disusun berdasarkan abjad serta memiliki keterangan yang berisi makna, pemakaian, atau terjemahan.

Kamus digital merupakan kamus yang dapat diakses melalui *gadget* dengan bantuan koneksi internet. Kamus Pintar Basdat berbasis *website* yang akan

dikembangkan meliputi pengertian istilah basis data khususnya materi tipe data.

### Basis Data

Menurut Begg & Connolly (2010:65), basis data merupakan suatu kumpulan data yang dirancang untuk menyimpan kebutuhan informasi dari sebuah organisasi secara logis. Pada pembelajaran SMK khususnya program kejuruan RPL, basis data merupakan ilmu yang mempelajari tentang hubungan antara kumpulan data yang disimpan tanpa *redundancy* atau perulangan yang menyebabkan data sama dengan yang lain untuk memenuhi berbagai kebutuhan aplikasi.

### Search Engine

*Search engine* merupakan program yang berfungsi untuk membantu pengguna mencari informasi di web dengan mengajukan pertanyaan atau menggunakan kata kunci (Williams & Sawyer, 2011:74).

Sistem pencarian saat ini banyak digunakan developer untuk memberikan fitur pada aplikasi agar lebih interaktif dan fleksibel untuk pengguna. Para pengguna aplikasi cukup mengisi kata yang ingin dicari kemudian aplikasi akan memberikan *feedback* berupa data atau dokumen yang mirip dengan kata yang diinputkan oleh pengguna.

### Bahasa Python

Menurut Swaroop (2013), *Python* ialah bahasa pemrograman yang memiliki karakter kuat serta mudah dipelajari. *Python* memiliki struktur data yang sederhana namun tingkat tinggi sehingga dapat dinilai efektif dan efisien untuk pemrograman berorientasi objek. *Python* memiliki sintaks yang dinamis. Sifat interpretasi *python* membuat bahasa pemrograman ini menjadi ideal dan dapat digunakan dalam pembuatan *script* dan pengembangan aplikasi di berbagai platform besar yang membutuhkan waktu singkat.

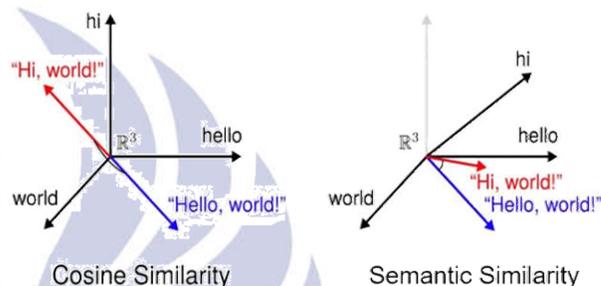
### Framework Flask

*Flask* merupakan *framework* yang mempunyai kerangka kerja yang luas dan memiliki beberapa layanan dasar. *Framework flask* mempunyai eksistensi dimana pengguna dapat bebas memilih sehingga dapat menghindari *bug* dan melakukan berbagai perintah kode program.

Flask memiliki 2 penemu utama, yaitu Werkzeug yang menciptakan *routing*, *debugging*, dan *Web Server Gateway Interface* (WSGI) dan template framework yang disediakan oleh Jinja. Werkzeug dan Jinja merupakan penemu utama pada pengembangan framework flask

### Algoritma Cosine Similarity

Algoritma *cosine similarity* yaitu sebuah metode dengan menggunakan dua buah kata yang menggunakan kata kunci atau *keywords* sebagai ukuran dari sebuah dokumen untuk menyatakan perhitungan kesamaan dua buah objek (Wahyuni dkk., 2017). Dalam implementasinya algoritma ini dapat dikembangkan menjadi algoritma *soft cosine similarity* dimana algoritma tersebut dapat mencari kedekatan kata berdasarkan sinonim kata. Algoritma *soft cosine similarity* memiliki 2 tahap yakni tahap *pre-processing data* dan tahap *processing data*



Gambar 1. Perbedaan *Cosine Similarity* dan *Semantic Similarity / Soft Cosine Similarity*.

Perbedaan algoritma *cosine similarity* dan *soft cosine similarity (semantic similarity)* dapat dilihat pada gambar 1. Proses *cosine similarity* dan *semantic similarity* dapat dijelaskan pada tahap *cosine similarity* kalimat "Hello, world!" dan "Hi, world!" memiliki arah vektor yang bertentangan. Namun, pada tahapan *semantic similarity* atau *soft cosine similarity* kalimat "Hi, world!" dengan "Hello, world!" memiliki arah vector berdekatan sehingga dianggap memiliki arti atau makna yang sama. Sehingga metode ini cocok digunakan dalam pengembangan kamus istilah basis data untuk mencari kesamaan makna atau istilah dalam setiap dokumen.

### Pendekatan Semantik

Pendekatan semantik yang merujuk pada algoritma *soft cosine similarity* pada umumnya sama seperti *cosine similarity* ditambah dengan relasi kesamaan makna antar kata. Fitur ini biasa digunakan untuk mencari kesamaan antar kata. Perbedaan yang mencolok dengan *cosine similarity* adalah *soft-cosine similarity* bisa melakukan penghitungan walaupun tidak ada kata yang sama antar variabel dengan persamaan di bawah atau sinonim kata bisa dihitung dekat oleh *soft-cosine similarity*.

### Dictionary Gensim

Untuk menghitung nilai kesamaan semantik digunakan Gensim yang merupakan library untuk digital

*document indexing* dan *similarity search* sebagai sarana dalam bahasa pemrograman *Python*. Dalam pemrosesan kata, kata dalam dataset harus diharuskan diubah menjadi id yang unik oleh gensim. Untuk merubah kata menjadi id yang unik, Gensim memungkinkan membuat objek Dictionary yang memetakan setiap kata ke id unik.

**Pendekatan Corpus dan Query**

*Gensim* memungkinkan sebuah sistem membaca teks atau kata dan memperbarui isi dari *dictionary*. Setiap dokumen pada *dictionary* yang telah diubah menjadi kata pada tahap *pre-processing* akan dikumpulkan dan disimpan pada *corpus* dalam bentuk *id* unik. *Corpus* merupakan kumpulan dokumen sebagai *bag of words* yang artinya setiap *corpus* berisi id setiap kata dan jumlah frekuensinya dalam dokumen tersebut. Sedangkan *query* merupakan dokumen yang berasal dari masukkan pengguna yang akan dicocokkan dengan dataset dan dihitung kedekatannya menggunakan *algoritma soft cosine similarity* yang akan dijalankan oleh *library* dari *gensim*.

**WordNet**

*WordNet* merupakan *database* leksikal *online* yang dirancang untuk digunakan di bawah kendali suatu program. Di dalamnya terdapat berbagai macam kosa kata benda, kata kerja, kata sifat serta kata keterangan yang diatur berdasarkan sinonimnya menurut sebuah konsep leksikal. Sebuah kata memiliki lebih dari satu pengertian dan dua kata tersebut memiliki kesamaan yang dapat dikatakan sinonim.

**METODE**

**Desain Penelitian**

Pengembangan kamus basis data berbasis website yaitu menggunakan metode pengembangan ADDIE (*Analysis, Design, Development, Implementation, Evaluation*). ADDIE merupakan proses umum yang digunakan oleh perancang instruksional dan pengembang pelatihan serta memiliki sifat dinamis dan fleksibel untuk membentuk pelatihan yang berguna dan sebagai unjuk alat dalam tampilan. ADDIE merupakan pendekatan yang menekankan bagaimana setiap komponen yang ada pada metode ini saling berinteraksi satu sama lain dan berkoordinasi sesuai dengan tahapan yang ada ( Sezer, 2013:137 ).

Pada metode ADDIE terdapat tahapan - tahapan dalam mengembangkan suatu produk yaitu :

- *Analysis*

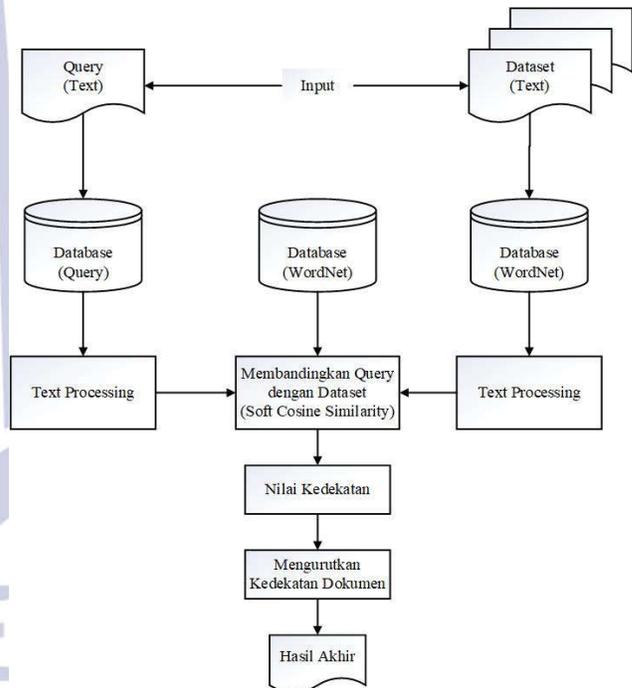
Kurangnya inovasi media pembelajaran membuat siswa kurang percaya diri pada proses pembelajaran. Selain itu media pembelajaran yang menggunakan search engine dengan metode *information retrival* sebagai teknologi canggih yang saat ini digunakan oleh banyak aplikasi

belum banyak diimplementasikan pada aplikasi pembelajaran atau *e-learning*.

Aplikasi kamus istilah basis data memerlukan sebuah *database* yang menyimpan sejumlah kosa kata dan maknanya mengenai basis data khususnya materi tipe data. Data tersebut akan diolah menggunakan algoritma *soft cosine similarity* untuk dicari kedekatannya berdasarkan *query* yang dimasukkan oleh pengguna.

- *Design*

Aplikasi kamus istilah basis data berbasis yang dikembangkan oleh penulis berjudul Pintar Basdat. Kamus Pintar Basdat berupa *search engine* dikembangkan menggunakan algoritma *soft - cosine similarity* sebagai pengembangan dari algoritma *cosine similarity* yang akan digambarkan pada rancangan sistem berikut.



Gambar 2. Cara kerja algoritma cosine similarity

Sedangkan desain tampilan *website* yang dirancang oleh penulis sebagai *interface* Kamus Pintar Basdat yaitu sebagai berikut



Gambar 3. Desain interface

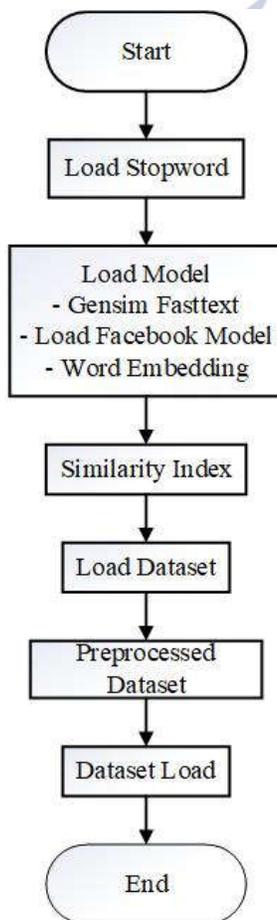
• *Development*

Berdasarkan rancangan sistem yang telah dibuat, penulis merancang *database* yang akan digunakan sebagai *dataset* dalam Dalam pemrosesan kata, penulis menggunakan *library gensim* yang dapat memproses kata menggunakan pendekatan *corpus* dan *query* dengan menghitung kedekatan kata berdasarkan perhitungan *similarity index*, *tf idf*, *similarity matrix* dan *soft cosine similarity*.

• *Implementation*

Proses implementasi algoritma *soft cosine similarity* memiliki 2 tahap yaitu tahap *pre-processing data* dan tahap *processing data*

1. Tahap *Pre – Processing Data*



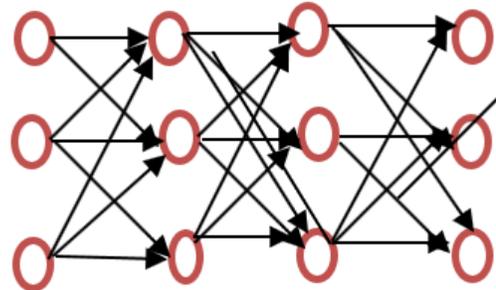
Gambar 4. Tahap *Pre-processing Data*

Tahap *pre-processing data* meliputi beberapa tahapan seperti pada gambar di atas dan dapat dijabarkan sebagai berikut :

1. *Load Model*

Sejumlah modul *gensim* digunakan dan dipersiapkan untuk pemrosesan dokumen dalam bentuk kata. Modul tersebut memiliki masing – masing fungsi yang akan

mendukung perhitungan algoritma *soft cosine similarity*, seperti *stopword* untuk memisahkan tanda baca, memisahkan imbuhan kata serta merubah huruf kapital menjadi huruf biasa. Selain itu terdapat juga modul *gensim fasttext* untuk membuat vector sebuah kata, *facebook model* seperti *word embedding* untuk penyematan kata yang memiliki representasi makna yang sama dari dataset dan database yang disediakan oleh *WordNet*. Proses representasi kata dapat dilihat pada gambar berikut.



Gambar 5. Proses representasi kata menggunakan *word2vec*

2. *Similarity Index*

Selanjutnya yaitu perhitungan kedekatan query dan kata pada *wordNet* menggunakan rumus *tf-idf*. Rumus *tf-idf* dapat dituliskan sebagai berikut.

$$w_{ij} = \frac{Tf_{if}}{\sum(Tf_{if})} \times \left( \log \left( \frac{N}{n_i} \right) + 1 \right)$$

Dimana :

- $w_{ij}$  = Bobot kata terhadap dokumen
- $Tf_{if}$  = Jumlah kemunculan kata/ *query*
- $N$  = Jumlah semua dokumen yang ada dalam *database*
- $n_i$  = Jumlah dokumen yang mengandung / *query*
- $t_j$  = (minimal terdapat satu kata yaitu term  $t_j$ )

3. *Load Dataset*

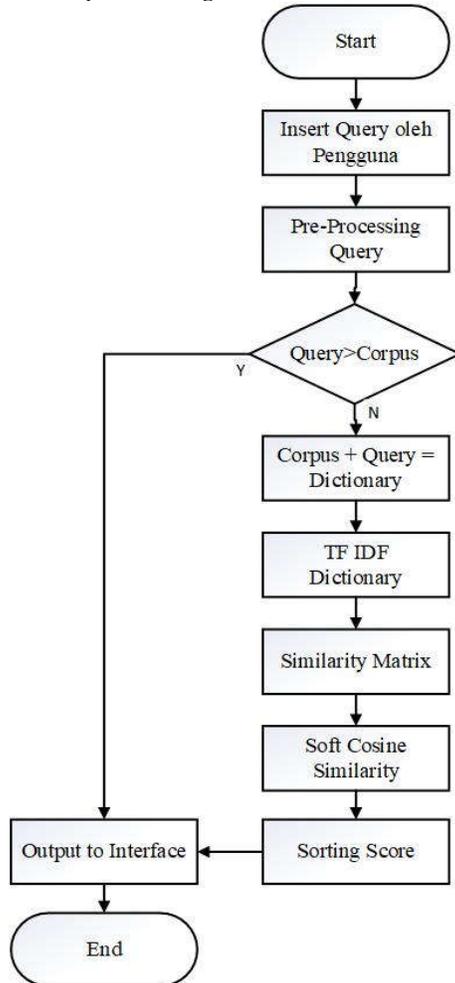
Kemudian pada proses ini, *dataset* yang berisi istilah – istilah basis data yang telah dibuat oleh peneliti dimasukkan ke dalam sistem untuk selanjutnya masuk ke tahap *pre-processing dataset*.

4. *Pre-processing Dataset*

Pada tahapan ini dilakukan proses *case folding*, *tokenizing*, dan *stopword*. *Case folding* bertujuan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Sedangkan *tokenizing* memisahkan kata berdasarkan spasi, enter, tabulasi, titik (.), dan koma (,). Dan *stopword* yaitu

penyimpanan kata yang memiliki frekuensi kemunculan tinggi namun tidak memiliki makna seperti “di”, “ke”, “dan”. Kata – kata yang disimpan ke dalam *stopword* biasanya berbentuk kata penghubung.

2. Tahap *Processing Data*



Gambar 6. Tahap *processing data*

Seperti gambar diatas, tahap *processing data* meliputi beberapa proses yaitu.

1. *Pre – Processing Query*

Pengguna memasukkan kata yang ingin dicari makna atau istilahnya. Kata yang dimasukkan pengguna disebut juga *query*. Setelah *query* masuk ke dalam sistem akan dilakukan *pre-processing* terhadap *query* yaitu berupa *case folding*, tokenisasi, dan *stopword*. Kemudian akan dilakukan pengecekan terhadap *query* karena *query* tidak dapat diproses jika kosong.

2. *Tf-idf Dictionary*

Kemudian *corpus* yang terdiri dari kata yang berasal dari *dataset* akan dijadikan satu dengan

*query* dan dimasukkan ke dalam dokumen dengan nama *dictionary*.

Selanjutnya dokumen dalam bentuk *dictionary* akan dihitung kedekatannya menggunakan algoritma *soft cosine similarity* melalui tahap *tf-idf* yaitu perhitungan kedekatan *query* dengan *dataset*.

3. *Similarity Matrix*

*similarity matrix* yaitu pembuatan *matrix* kata yang berdekatan kemudian perhitungan *soft-cosine* menggunakan *library gensim*

4. *Soft Cosine Similarity*

Tahap terakhir yakni perhitungan kedekatan istilah atau makna kata dengan dokumen menggunakan perhitungan *soft cosine* oleh *word2vec* dan menghasilkan skor yang telah dihasilkan akan diurutkan berdasarkan yang terbesar yaitu menunjukkan dokumen dari *dataset* yang paling mendekati makna *query* yang dimasukkan pengguna.

Rancangan penelitian yang digunakan yaitu perancangan *one group pre-test post-test*. *One group pre-test post-test design* membandingkan keadaan sebelum dan sesudah diberi perlakuan. *Pre-test* untuk sebelum perlakuan dan *post-test* sesudah perlakuan dengan lebih akurat (Sugiyono, 2015: 74 ). Dalam rancangan penelitian ini, sample diberikan kepada siswa ketika sebelum diberi perlakuan, kemudian siswa diberikan perlakuan berupa penggunaan kamus istilah basis data sebagai media pembelajaran. Kemudian siswa diberi *post-test*. *Pre-test* dan *post-test* berupa soal – soal mengenai materi tipe data dalam basis data.

$$O_1 \times O_2$$

Keterangan:

O1 : Observasi pertama untuk mengetahui kemampuan siswa dalam memahami istilah basis data dan materi tipe data sebelum mendapat perlakuan.

X : Perlakuan berupa penggunaan kamus istilah basis data sebagai media pembelajaran

O2 : Observasi akhir untuk mengetahui kemampuan siswa dalam memahami istilah basis data dan materi tipe data setelah mendapat perlakuan.

Sedangkan prosedur penelitian dapat digambarkan sebagai berikut



Gambar 7. Prosedur Penelitian

Berdasarkan gambar prosedur penelitian diatas, Observasi Pertama, digunakan untuk mengamati kemampuan pemahaman materi tipe data berdasarkan instrument soal yang telah diberikan. Pada tahap eksperimen, diberikan perlakuan dengan pre – experimental design yaitu pembelajaran menggunakan media kamus basis data berbasis website. Observasi Akhir berupa pengukuran akhir kemampuan pemahaman materi tipe data setelah siswa diberikan perlakuan pembelajaran menggunakan kamus pintar basdat Tahapan proses tersebut digunakan untuk mengetahui pengaruh media terhadap pemahaman materi pada siswa.

### Sampel Penelitian

Sampel harus diperhatikan dalam teknik pengambilan data meskipun kualitas penelitian tidak selalu ditentukan oleh besarnya sampel, namun berdasarkan dasar teori, desain penelitian, serta kualitas pelaksanaan dan pengolahan penelitian ( Hadi&Akdon, 2005). Berdasarkan pernyataan tersebut, peneliti menjadikan sebagian dari populasi yaitu siswa kelas XI RPL SMK Negeri 1 Surabaya menjadi sampel penelitian.

### Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan yakni melalui observasi, kuesioner dan instrument soal. Observasi digunakan untuk mengamati pembelajaran , penggunaan media pembelajaran dan metode pengajaran saat proses pembelajaran. Kuesioner berisi daftar pertanyaan yang berhubungan dengan penggunaan media pembelajaran kamus istilah basis data pada proses pembelajaran.

Sedangkan instrumen soal berisi beberapa pertanyaan mengenai materi tipe data pada basis data. Terdapat dua instrument soal yaitu pre-test dan post-test. Pre – test dikerjakan oleh siswa sebelum siswa diberi perlakuan, sedangkan post-test dikerjakan oleh siswa seteah siswa diberi perlakuan bersamaan dengan kuesioner penggunaan media pembelajaran. masing – masing intrumen soal dan kuesioner telah divalidasi oleh para ahli sebelum melaukan penelitian.

### Teknik Analisis Data

Untuk menganalisis data, pada penelitian ini peneliti menggunakan analisis kelayakan media, RPP, materi, kuesioner, dan instrumen soal. Analisis data instrumen penelitian menggunakan perhitungan presentase kelayakan berdasarkan hasil penilaian dari para ahli. Kemudian analisis hasil belajar dilakukan dengan adanya uji hipotesis menggunakan metode *Shapiro-wilk* dan uji normalitas menggunakan metode *Paired T-Test* dengan bantuan aplikasi IBM SPSS *Statistics* 25 untuk

mengetahui adanya perbedaan hasil belajar setelah siswa diberi perlakuan.

### HASIL DAN PEMBAHASAN

Hasil pengembangan aplikasi kamus istilah basis data menggunakan algoritma cosine similarity dengan menerapkan soft-cosine diberi nama “Kamus Pintar Basdat” yaitu media yang dapat menampilkan istilah atau makna kata berdasarkan kedekatan maknanya atau sinonim dari kata yang dimasukkan oleh pengguna. Kamus Pintar Basdat dapat diakses dengan cukup mudah melalui *link* dan jaringan internet.

Berikut ini merupakan tampilan dari *website* Kamus Pintar Basdat :



Gambar 8. Tampilan Awal

Kamus Pintar Basdat memiliki satu *interface* saja yang menampilkan judul, prosedur penggunaan, masukkan kata dan tombol search disampingnya. Sebagai contoh pengguna memasukkan kata “pecahan” kemudian akan muncul hasilnya seperti gambar di bawah ini. kata yang muncul yaitu *double* dan *float* karena 2 kata tersebut memiliki kedekatan atau makna yang sama dengan kata”pecahan”.



Gambar 9. Tampilan Hasil

### Hasil Analisis Validasi Data

Intrumen penelitian berupa soal pretest-posttest dan angket respon siswa, perangkat pembelajaran serta media pembelajaran yang digunakan pada penelitian ini telah melalui tahap validasi oleh 2 dosen dari Jurusan Teknik Informatika UNESA dan satu guru XI RPL di SMKN 1 Surabaya. Berikut bidang keahlian dari validator yang melakukan validasi instrument penelitian ini.

Tabel 1. Daftar Nama Validator

No.	Bidang Keahlian	Keterangan
1.	Pendidikan Teknologi Informasi	Dosen UNESA
2.	Teknologi Informasi	Dosen UNESA
3.	Pendidikan Vokasi	Guru SMK

Hasil analisis kelayakan RPP, materi, angket respon siswa, instrumen soal dan media dijabarkan sebagai berikut.

Tabel 2. Hasil Analisis Validasi Ahli

No.	Instrumen Penelitian	Presentase	Kriteria
1.	RPP	86,3%	Sangat Layak
2.	Materi Pembelajaran	84,8%	Sangat Layak
3.	Angket Respon Siswa	89,2%	Sangat Layak
4.	Instrumen Soal	85%	Sangat Layak
5.	Media “Kamus Pintar Basdat”	80,5%	Layak

**Pembahasan Hasil Penelitian**

Penelitian dilakukan di SMKN 1 Surabaya dengan total siswa sebagai sampel penelitian yaitu 35 siswa kelas XI RPL. Tahap pertama peneliti melakukan observasi kemampuan siswa melalui pre-test. Kemudian proses pembelajaran dilakukan melalui google classroom dan pertemuan melalui google meet sehingga peneliti dapat melakukan pengamatan terhadap siswa. Selanjutnya yaitu tahap uji coba produk, siswa diberi kesempatan untuk belajar dengan Kamus Pintar Basdat. Pada tahap akhir, siswa melakukan pengerjaan post-test sebagai alat ukur kemampuan siswa setelah belajar menggunakan media pembelajaran.

Kemudian peneliti menggunakan Uji *Paired Sample T-Test* untuk mengetahui pengaruh hasil belajar siswa terhadap penggunaan media pembelajaran. Peneliti menggunakan dimana sebelumnya data nilai siswa perlu diuji normalitasnya menggunakan metode *Shapiro-wilk*. Penulis juga melakukan analisis jawaban angket respon siswa yang digunakan untuk mengetahui respon dan motivasi belajar siswa menggunakan media Kamus Pintar Basdat.

1. Analisis Hasil Belajar

a. Uji Normalitas

Uji normalitas bertujuan untuk mengetahui data berdistribusi normal atau tidak. Uji normalitas sebagai syarat untuk melakukan uji T. berikut merupakan tabel hasil uji normalitas data nilai siswa kelas XI RPL

Tabel 3. Hasil Uji Normalitas (*Shapiro-wilk*)

<i>Tests of Normality</i>			
	Statistic	df	Sig.
Nilai Pre-test	.977	35	.651
Nilai Post-test	.941	35	.059

Dari table diatas, data telah diuji menggunakan metode *Shapiro-wilk* dimana hasil signifikansi pada nilai pre-test yaitu  $0.651 > 0.05$  menunjukkan bahwa nilai pre-test berdistribusi normal. Sedangkan hasil uji nilai post-test memperoleh nilai  $0.059 > 0.05$  sehingga disimpulkan bahwa nilai post-test berdistribusi normal.

b. Uji *Paired Sample T-Test*

Berdasarkan hasil uji normalitas diperoleh hasil bahwa data berdistribusi normal, maka data pre-test dan post-test akan diuji dengan *Paired Sample T-Test* untuk mengetahui pengaruh hasil belajar siswa sebelum dan setelah menggunakan media Kamus Pintar Basdat. Berikut tabel hasil uji *Paired Sample T-Test* pada data nilai siswa kelas XI RPL.

Tabel 4. Hasil Uji Statistik Data

<i>Paired Samples Statistics</i>					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-test	79.4286	35	9.11700	1.54105
	Post-test	87.7857	35	7.73579	1.30759

Tabel 5. Hasil Uji *Paired Sample T-Test*

<i>Paired Samples Test</i>					
		Mean	t	df	Sig. (2-tailed)
Pair 1	Pre_test – Post test	-8.35714	-9.892	35	.000

Berdasarkan tabel 4, rata-rata nilai *pre-test* dan *post-test* menunjukkan peningkatan dari yang sebelumnya 79,42 menjadi 87,7. Kemudian dapat dilihat pada tabel 5, nilai signifikansi yang diperoleh yaitu  $0,000 < 0,05$  yang menunjukkan bahwa ada perbedaan antara hasil belajar sebelum dan sesudah Kamus Pintar Basdat. Sehingga dari analisis nilai signifikansi dan rata – rata nilai siswa pada tabel hasil uji *Paired Sample T-Test* dapat diketahui bahwa media pembelajaran Kamus Pintar Basdat dapat meningkatkan hasil belajar siswa kelas XI RPL.

2. Analisis Angket Respon Siswa

Hasil pengisian angket respon siswa yang didapatkan dari siswa kelas XI RPL digunakan untuk mengetahui bagaimana respon siswa dan motivasi belajar siswa dalam pembelajaran

menggunakan Kamus Pintar Basdat. Angket respon siswa mempunyai sub variable yang diukur yaitu respon terhadap materi, media dan pengoperasiannya serta motivasi belajar. Hasil analisis respon siswa yaitu memperoleh presentase 80,6% dengan kategori respon sangat positif setelah menggunakan media Kamus Pintar Basdat.

## SIMPULAN

1. Berdasarkan hasil pengembangan aplikasi kamus pintar basdat berbasis web dengan menggunakan algoritma cosine similarity dan menerapkan metode soft-cosine yang digunakan untuk proses pembelajaran pada siswa kelas XI RPL telah dilaksanakan sesuai dengan prosedur penelitian. Pengembangan aplikasi pada penelitian ini menggunakan metode ADDIE.
2. Hasil uji kelayakan media pembelajaran kamus pintar basdat mendapatkan presentase 80,5% dengan kategori layak dan dapat disimpulkan bahwa media tersebut dapat digunakan pada proses pembelajaran.
3. Kemudian hasil uji instrumen penelitian yang telah divalidasi oleh ahli yaitu hasil uji kelayakan RPP dikategorikan sangat layak dengan presentase 86,7%, presentase uji kelayakan materi yaitu 81,4% dengan kategori sangat layak, dan uji kelayakan angket respon siswa yaitu 86,4% dengan kategori sangat layak serta hasil uji kelayakan instrument soal yaitu 84,3% dengan kategori sangat layak.
4. Hasil uji *Paired Sample T-Test* yaitu  $0,000 < 0,05$  yang menunjukkan bahwa ada signifikansi antara hasil belajar sebelum dan sesudah menggunakan Kamus Pintar Basdat. Sehingga disimpulkan bahwa penggunaan media pembelajaran Kamus Pintar Basdat pada pembelajaran materi tipe data, mempengaruhi hasil belajar siswa menjadi lebih tinggi dan siswa memiliki motivasi belajar yang lebih. Hal ini dapat ditunjukkan pada hasil nilai signifikansi pada Kemudian rata – rata hasil belajar siswa menunjukkan peningkatan dari nilai pre-test dan post-test. Selain itu respon siswa menunjukkan sangat positif dengan presentase 80,6% dari hasil analisis angket yang telah diisi-oleh siswa.

## SARAN

1. Pengembangan Kamus Pintar Basdat menggunakan algoritma cosine similarity perlu dikembangkan lebih lanjut seiring dengan

perkembangan teknologi untuk mendukung proses belajar siswa.

2. Kamus yang dikembangkan oleh penulis masih terbatas materi pembelajaran sehingga penelitian ini masih ada celah untuk dikembangkan dengan cara memperluas sasaran pembelajaran sehingga cakupan materi menjadi lebih banyak.
3. Dalam implementasi penggunaan Kamus Pintar Basdat perlu dikelola lebih lanjut perihal jaringan atau server yang digunakan untuk aplikasi sehingga *load* server dapat berjalan lebih optimal.

## UCAPAN TERIMA KASIH

Dalam hal ini, Penulis mengucapkan terima kasih kepada :

1. Ayah dan Ibu yang selalu berdoa, berjuang, dan tak pernah putus asa untuk anak terakhirnya
2. Dosen pembimbing skripsi selaku dosen pembimbing skripsi Ibu Yeni Anistiyasari, S.Pd., M.Kom. yang telah bersedia memberikan arahan, meluangkan waktu serta membimbing dengan sangat baik dari awal perkuliahan hingga masa akhir perkuliahan.
3. Keluarga kedua PTI 2017 yang selalu mendukung dan memberikan arahan bersama – sama.

## DAFTAR PUSTAKA

- 99, G. (2021, Juni 8). *Word Embedding Tutorial: Word2vec with Gensim [EXAMPLE]*. Retrieved from Guru 99: <https://www.guru99.com/word-embedding-word2vec.html>
- Chang, Y.-H., & dkk. (2018). Cognitive resources allocation in computermediated dictionary asisted learning: From word meaning to inferential comprehension,. *Computers & Education*. doi:doi: 10.1016/j.compedu.2018.08.013.
- Grinberg, & Miguel. (2014). *Flask Web Development*. United States of America: O'Reilly Media, Inc.
- Ismail, N. S., & dkk. (2017). The Effect of Mobile Problem-Based Science Dictionary Aplication DicScience PBL Towards Students' Mastery of Scientific Terms and Critical Thinking. doi:10.1109/LaTiCE.2017.19.
- Prabhakaran, S. (2021, April 16). *Cosine Similarity*. Retrieved from *Cosine Similarity – Understanding the math and how it works (with python codes)*: <https://www.machinelearningplus.com/nlp/cosine-similarity/>
- Prabhakaran, S. (2021, May 25). *Cosine Similarity – Understanding the math and how it works (with python codes)*. Retrieved from *Machine Learning +:* <https://www.machinelearningplus.com/nlp/cosine-similarity/>

Pramudya, F. Y., & dkk. (2019). Pemberian Peringkat Komentar pada Community Question Answering dengan Fitur Soft-Cosine Semantic Similarity untuk Kasus Question-External Coment. e-Proceeding of Engineering : Vol.6, No.1, 2199.

Rayanto, Y. H., & Sugianti. (2020). Penelitian Pengembangan Model ADDIE dan R2D2 : Teori dan Praktek. Pasuruan: Lembaga Academic & Research Institute.

Salma, D. (2013). Mozaik Teknologi Pendidikan: E-Learning. Jakarta: Prenada Media Group.

Sidorov, G., & dkk. (2014). Soft Similarity and Soft Cosine Measure: Similarity of Feature in Vector Space Model. *Computacion y Sistemas* Vol. 18, No. 3, 491-504. doi:10.13053/CyS-18-3-2043

Sugiyono. (2017). Metode Penelitian Kuantitatif, Kuallitatif, dan R&D. Bandung: Alfabeta.

Swaroop. (2020, Maret 01). A byte of Phyton. Retrieved from [https://www.ibiblio.org/swaroopch/byteofpython/files/120/byteofpython\\_120.pdf](https://www.ibiblio.org/swaroopch/byteofpython/files/120/byteofpython_120.pdf): <http://www.swaroopch.com/notes/Python>

Tsuruoka, Y., & Tsuiji, J. (2004). Improving the performance of dictionary-based aproaches in protein name recognition. *Journal of Biomedical Informatics*, 37 (2004), 461–470. doi:<https://doi.org/10.1016/j.jbi.2004.08.003>

Wahdiat, M. T., & Ade Romadhony, S. (2019). Analisis Hasil Penerapan Metode Distributional Semantic untuk Kesaman Semantik pada Bahasa Indonesia. e-Proceding of Engineer (99, 2021)ing : Vol.6, No.2, 8482.

Wahyuni, R. T., & dkk. (2017). Penerapan Algoritma Cosine Simillarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro* Vol. 9 No. 1. doi:<https://doi.org/10.15294/jte.v9i1.10955>

Zwart, D. P., & dkk. (2020). Effects of Digital Learning Materials on nursing students' matematics learning, self-efficacy, and task value in vocational education. *Nurse Education Practice*,, 44 102755. doi:10.1016/j.nepr.2020.102755