

**YAYASAN LEMBAGA PENDIDIKAN ISLAM DAERAH RIAU
UNIVERSITAS ISLAM RIAU
FAKULTAS TEKNIK**

**KLASIFIKASI KUALITAS RETENSI SURFAKTAN
PADA BATUAN SANDSTONE DENGAN MENGGUNAKAN
METODE *K-NEAREST NEIGHBOR***

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Penyusunan Skripsi Pada Fakultas Teknik
Universitas Islam Riau Pekanbaru



SRI REZEKI
153510470

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM RIAU
PEKANBARU
2020**

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Dengan mengucapkan Alhamdulillah Robbil 'Alamin, berkat rahmat dan hidayah Allah SWT serta nikmat yang tak terhingga, penulis dapat menyelesaikan laporan skripsi ini dengan judul "KLASIFIKASI KUALITAS RETENSI SURFAKTAN PADA BATUAN SANDSTONE DENGAN MENGGUNAKAN METODE *K-NEAREST NEIGBOR*" sebagai salah satu syarat untuk memperoleh gelar sarjana teknik pada Fakultas Teknik Program Studi Teknik Informatika Universitas Islam Riau.

Dalam penyusunan laporan skripsi ini, penulis menyadari banyak mendapat hambatan dan tantangan. Namun, dalam penyelesaian penulisan ini penulis tidak terlepas dan terwujud tanpa bimbingan, pengarahan, saran dan bantuan moril maupun non-moril dari berbagai pihak. Untuk itu, dalam kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada yang terhormat :

1. Bapak Dr. Arbi Haza Nasution, B.IT.(Hons), M.IT selaku ketua Program Studi Teknik Informatika atas bimbingan dan motivasi yang telah diberikan yang telah membantu dalam kelancaran penyelesaian laporan skripsi ini.
2. Ibu Ir. Des Suryani, M.Sc selaku dosen pembimbing skripsi yang telah ikhlas dan sabar memberikan bimbingan, saran dan motivasi yang bermanfaat dalam penyusunan laporan skripsi.

3. Seluruh Dosen Teknik Informatika yang telah banyak memberikan ilmu akademik maupun non-akademik selama berada di bangku pendidikan.
4. Teristimewa untuk kedua orang tua tercinta, abang dan adik-adik yang senantiasa selalu mendoakan, serta memberikan dukungan baik moril maupun materil dalam menyelesaikan laporan skripsi ini.
5. Rekan-rekan mahasiswa/i khususnya kelas D angkatan 2015 serta semua pihak yang telah memberikan bantuan, semangat dan motivasi selama penyusunan laporan skripsi ini.

Penulis menyadari bahwa dalam penyusunan laporan skripsi ini masih banyak kekurangan, oleh karena itu penulis mengharapkan adanya kritik dan saran yang bersifat membangun untuk memperbaiki laporan skripsi ini.

Akhir kata penulis berharap penyusunan laporan skripsi ini dapat bermanfaat bagi semua pihak yang membacanya dan dapat dikembangkan lebih lanjut, Amin.

Wassalammu'alaikum. Warahmatullahi Wabarakatuh

Pekanbaru, Februari 2021

Sri Rezeki

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan *Sandstone* dengan Menggunakan Metode *K Nearest Neighbor*

Sri Rezeki
Fakultas Teknik
Program Studi Teknik Informatika
Universitas Islam Riau
Email : rezekisri638@student.uir.ac.id

ABSTRAK

Batuan *Sandstone* merupakan salah satu jenis batuan sedimen yang terbentuk dari sementasi dari butiran-butiran pasir yang terbawa oleh aliran sungai, angin, dan ombak dan akhirnya terakumulasi pada suatu tempat yang didalam batuan tersebut mengandung banyak minyak. Namun untuk mengeluarkan minyak tersebut dapat dilakukan dengan cara menginjeksikan larutan surfaktan. Surfaktan bekerja dengan cara menurunkan *interfacial tension* antara minyak dan batuan, hingga minyak yang terperangkap tersebut dapat terlepas dari batuan, Kejadian tersebut dinamakan retensi surfaktan yang merupakan permasalahan utama pada saat penginjeksian surfaktan pada batuan. Adapun parameter yang mempengaruhi retensi surfaktan yaitu *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant*. Untuk menyelesaikan masalah tersebut, penelitian ini membuat suatu sistem klasifikasi untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* dengan menggunakan metode *K Nearest Neighbor*. Metode ini bekerja dengan mencari kedekatan jarak suatu data dengan dengan data lain. Hasil dari pengujian nilai K terbaik diperoleh pada K=3 dan hasil pengujian akurasi sistem klasifikasi kualitas retensi surfaktan adalah 82,22% dan hasil implementasi yang didapat sistem sebesar 76,2% menunjukkan bahwa sistem klasifikasi surfaktan pada batuan *sandstone* ini dapat diterapkan.

Kata Kunci : Data Mining, Klasifikasi, *K Nearest Neighbor*, *Sandstone*

Surfactant Retention Quality Classification On Sandstone Rocks Using the Method K Nearest Neighbor

Sri Rezeki
Fakultas Teknik
Program Studi Teknik Informatika
Universitas Islam Riau
Email : rezekisri638@student.uir.ac.id

ABSTRACT

Sandstone is a type of sedimentary rock formed from cementation of grains of sand carried by rivers, wind, and waves and eventually accumulates in a place where the rock contains a lot of oil. However, to remove the oil can be done by injecting a surfactant solution. Surfactants work by reducing the interfacial tension between oil and rock, so that the trapped oil can be released from the rock. This event is called surfactant retention which is the main problem when injecting surfactants into rocks. The parameters that affect surfactant retention are Absolute Permeability, Total Acid Number, Temperature, Co-solvent, Polimer Salinity, Max Effluent pH, Mobility Ratio and Molecular weight (MW) Surfactant. To solve this problem, this study makes a classification system to determine the quality of surfactant retention in sandstone using the K Nearest Neighbor method. This method works by finding the proximity of a data to other data. The results of testing the best K value obtained at $K = 3$ and the results of testing the accuracy of the surfactant retention quality classification system is 82.22% and the implementation results obtained by the system are 76.2% indicating that the surfactant classification system in sandstone rocks can be applied.

Keywords: *Data Mining, Classification, Method K Nearest Neighbor, Sandstone*

DAFTAR ISI

	Hal
KATA PENGANTAR	ii
ABSTRAK	iv
ABSTRACT	v
DAFTAR ISI	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	4
1.4 Batasan Masalah	4
1.5 Tujuan	5
1.6 Manfaat Penelitian	5
BAB II LANDASAN TEORI	6
2.1 Tinjauan Pustaka	6
2.2 Dasar Teori	8
2.2.1 <i>Enhanced Oil Recovery</i> (EOR)	8
2.2.2 <i>Chemical Flooding</i>	9
2.2.3 Surfaktan.....	14
2.2.4 Data Mining	16

2.2.5	Klasifikasi	17
2.2.6	Metode <i>k-Nearest Neighbour</i>	20
2.2.7	<i>K Fold Cross Validation</i>	23
2.3.	Tool dalam Analisa dan Perancangan Sistem	25
2.3.1	Use Case	25
2.3.2	<i>Data Flow Diagram</i>	26
2.3.3	<i>Context Diagram</i>	27
2.3.4	<i>Entity Relationship Diagram</i>	29
2.3.5	Program <i>Flowcart</i>	30
2.3.6	<i>Structured Query Language</i>	31
2.4	Pemrograman	32
2.4.1	<i>Hypertext Preprocessor</i>	32
2.4.2	HTML	33
2.5	Pengukuran Akurasi	34
BAB III	METODOLOGI PENELITIAN	35
3.1	Alat Penelitian yang Digunakan	35
3.2	Bahan Penelitian yang Digunakan	35
3.3	Analisis Sistem yang Sedang Berjalan	36
3.4	Perancangan Sistem	37
3.4.1	<i>Context Diagram</i>	37
3.4.2	<i>Hierarchy Chart</i>	38
3.4.3	<i>Data Flow Diagram (DFD) Level 0</i>	39
3.4.4	<i>Data Flow Diagram (DFD) Level 1 Proses 1</i>	40

3.4.5 <i>Data Flow Diagram (DFD) Level 1 Proses 2</i>	41
3.5 Rancangan Desain	41
3.5.1 <i>Desain Output</i>	41
3.5.2 <i>Desain Input</i>	44
3.5.3 <i>Desain Database</i>	47
3.6 Pehitungan Manual	50
3.7 Desain Logika Program	60
BAB IV Hasil dan Pembahasan	66
4.1 Pengujian Hasil	66
4.1.1 Pengujian Akurasi Sistem.....	66
4.2 Pengujian <i>BlackBox</i>	68
4.2.1 Pengujian <i>Login</i>	68
4.2.2 Pengujian Data <i>User</i>	70
4.2.3 Pengujian Data <i>Training</i>	71
4.2.4 Pengujian Data <i>Testing</i>	74
4.2.5 Pengujian <i>K Fold Cross Validation</i>	77
4.3 Pengujian Sistem Terhadap Pengguna.....	78
BAB V Kesimpulan dan Saran	83
5.1 Kesimpulan	83
5.2 Saran	83
DAFTAR PUSTAKA	84
LAMPIRAN	86

DAFTAR TABEL

	Hal
Tabel 2.1 Pengelompokan Retensi Surfaktan	40
Tabel 2.2 Simbol <i>Use Case Diagram</i>	41
Tabel 2.3 Notasi-Notasi pada DFD Edward Yourdon dan Tom DeMarco	42
Tabel 2.4 Simbol-Simbol ERD dengan Notasi Chen	43
Tabel 2.5 Simbol dan Fungsi <i>Flowchart</i>	52
Tabel 3.1 Tabel <i>User</i>	55
Tabel 3.2 Tabel <i>Training</i>	57
Tabel 3.3 Tabel <i>Testing</i>	60
Tabel 3.4 Tabel K Optimal	63
Tabel 3.5 Tabel <i>Data Training</i>	50
Tabel 3.6 Tabel <i>Data Testing</i>	51
Tabel 3.7 Tabel Perhitungan Jarak	54
Tabel 3.8 Tabel Pengurutan Data	55
Tabel 4.1 Pengujian Akurasi Sistem	52
Tabel 4.2 Tabel <i>Confusion Matrix</i>	55
Tabel 4.3 Tabel Pengujian <i>Form Login</i>	57
Tabel 4.4 Tabel Pengujian <i>Data User</i>	60
Tabel 4.5 Tabel Pengujian <i>Data Training</i>	63
Tabel 4.6 Tabel Pengujian <i>Data Testing</i>	50
Tabel 4.7 Tabel Pengujian <i>K Fold Cross Validation</i>	51
Tabel 4.8 Tabel Hasil Nilai Persentase Tiap Pertanyaan Kuisisioner	54

DAFTAR GAMBAR

	Hal
Gambar 2.1 Contoh Klasifikasi pada <i>k-Nearest Neighbor</i>	14
Gambar 2.2 Pembagian data pada metode <i>10-Fold Cross Validation</i>	15
Gambar 3.1 <i>Use Case</i> Sistem yang sedang berjalan	15
Gambar 3.2 <i>Context Diagram</i> Sistem Penentuan Retensi Surfaktan	17
Gambar 3.3 <i>Hierarchy Chart</i> Klasifikasi Retensi Surfaktan	18
Gambar 3.4 <i>Data Flow Diagram</i> (DFD) Level 0	18
Gambar 3.5 <i>Data Flow Diagram</i> (DFD) Level 1 Proses 1	18
Gambar 3.6 <i>Data Flow Diagram</i> (DFD) Level 1 Proses 2	19
Gambar 3.7 Desain <i>Output Data Training</i>	19
Gambar 3.8 Desain <i>Output Data Testing</i>	19
Gambar 3.9 Desain <i>Output</i> Perhitungan <i>k-NN</i>	19
Gambar 3.10 Desain <i>Output</i> Nilai <i>K</i> Optimal dan Akurasi.....	20
Gambar 3.11 Desain <i>Input Login</i>	20
Gambar 3.12 Desain <i>Input Data Training</i>	20
Gambar 3.13 Desain <i>Input Data Training</i>	20
Gambar 3.14 Desain <i>Input</i> Perhitungan <i>K-Fold Cross Validation</i>	21
Gambar 3.15 Program <i>Flowchart</i> Login	21
Gambar 3.16 Program <i>Flowchart</i> Menu Utama	21
Gambar 3.17 Program <i>Flowchart</i> <i>Input Data Training</i>	21
Gambar 3.18 Program <i>Flowchart</i> Perhitungan <i>Data Testing</i>	21
Gambar 3.19 Program <i>Flowchart</i> <i>K Fold Cross Validation</i>	21

Gambar 4.1 Tampilan <i>Login</i>	21
Gambar 4.2 Tampilan Tambah Data <i>User</i>	21
Gambar 4.3 Tampilan Edit Data <i>User</i>	21
Gambar 4.4 Tampilan Menu Data <i>Training</i>	21
Gambar 4.5 Tampilan Tambah Data <i>Training</i>	21
Gambar 4.6 Tampilan Berhasil Tambah Data <i>Training</i>	21
Gambar 4.7 Tampilan Edit Data <i>Training</i>	21
Gambar 4.8 Tampilan Menu Data <i>Testing</i>	21
Gambar 4.9 Tampilan Tambah Data <i>Testing</i>	21
Gambar 4.10 Tampilan Berhasil Tambah Data <i>Testing</i>	21
Gambar 4.11 Tampilan Edit Data <i>Testing</i>	21
Gambar 4.12 Pengujian Data <i>Testing</i>	21
Gambar 4.13 Pengujian <i>input</i> nilai <i>k</i>	21
Gambar 4.14 Pengujian <i>k Fold Cross Validation</i>	21
Gambar 4.15 Grafik Hasil Kuisisioner Pengujian Sistem Terhadap Pengguna .	21

DAFTAR LAMPIRAN

	Hal
Lampiran 3.1 <i>Data Training</i>	54
Lampiran 3.2 <i>Data Testing</i>	57



BAB I

PENDAHULUAN

1.1. Latar Belakang

Minyak bumi merupakan sumber energi yang masih menjadi primadona dan belum tergantikan oleh bahan lain hingga saat ini. Permintaan dunia terhadap minyak bumi diperkirakan akan terus meningkat sedangkan produksi minyak bumi dunia terus menurun dari tahun ke tahun. Faktor utama penyebab penurunan produksi minyak bumi di Indonesia adalah sumur-sumur minyak di Indonesia merupakan ladang minyak tua. Banyak cara yang dapat dilakukan dalam meningkatkan *recovery* minyak seperti injeksi air ditambahkan dengan bahan kimia guna meningkatkan energi dorong pada *reservoir* melalui pendesakan minyak di antara batuan sehingga membentuk kondisi yang baik untuk memaksimalkan *recovery* minyak dimana proses tersebut dikenal dengan proses *Enhanced Oil Recovery* (EOR).

Enhanced Oil Recovery (EOR) merupakan teknik lanjutan untuk mengangkat minyak jika berbagai teknik dasar sudah dilakukan tetapi hasilnya tidak ekonomis. *Chemical Flooding* adalah salah satu metode EOR terbaik yang mampu memulihkan 35% dari *Original Oil in Place* (OOIP) (Mandal, 2015). *Chemical Flooding* adalah proses penginjeksian bahan kimia yang fungsinya dapat meningkatkan perolehan minyak. *Chemical Flooding* terdiri dari beberapa metode, salah satunya *Surfactant Flooding*. Injeksi surfaktan ini ditujukan untuk memproduksi *residual oil* yang ditinggalkan oleh *water drive*, dimana minyak

yang terjebak oleh tekanan kapiler tersebut tidak dapat bergerak. Minyak tersebut dapat dikeluarkan dengan menginjeksikan larutan surfaktan. Surfaktan bekerja dengan cara menurunkan *interfacial tension* antara minyak dan batuan, hingga minyak yang terperangkap tersebut dapat terlepas dari batuan. surfaktan mudah terserap pada permukaan (matriks) batuan . Kejadian tersebut dinamakan retensi surfaktan yang merupakan permasalahan utama pada saat penginjeksian surfaktan pada batuan. Salah satu batuan yang sering dilakukan injeksi surfaktan adalah batuan *sandstone*. Batuan *Sandstone* atau yang biasa disebut dengan batu pasir merupakan salah satu jenis batuan sedimen yang terbentuk dari sementasi dari butiran-butiran pasir yang terbawa oleh aliran sungai, angin, dan ombak dan akhirnya terakumulasi pada suatu tempat yang didalam batuan tersebut mengandung banyak minyak

Sistem yang dibangun pada penelitian ini merupakan sistem yang menerapkan data mining. Data mining adalah proses menemukan wawasan,pola menarik, serta deskriptif, dimengerti dan prediksi dari model data dari data berskala besar. Dengan ini melihat sifat dasar data yang dimodelkan sebagai matriks data, yang menekankan pandangan geometris dan aljabar, serta interpretasi probabilistik dari data (Mohammed J. Zaki and Wagner Meira, 2014). Adapun metode klasifikasi yang digunakan dalam sistem ini adalah *k-Nearest Neighbor*, yaitu algoritma pengklasifikasi data sederhana dimana perhitungan jarak terpendek dijadikan ukuran untuk mengklasifikasi suatu kasus baru berdasarkan ukuran kemiripan (Pandie 2012).

Berdasarkan hasil penelitian teknik perminyakan pada Kamari *et al* (2015) berpendapat bahwa nilai dari retensi surfaktan dipengaruhi oleh beberapa parameter, yaitu: *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight* (MW) *Surfactant* dan kualitas retensi surfaktan sebagai outputnya. Dalam menentukan kualitas retensi surfaktan mahasiswa, dosen, bahkan peneliti masih merasa kesulitan sebab tidak adanya aplikasi/*software* yang digunakan untuk mempermudah pekerjaan pengguna.

Berdasarkan dari permasalahan di atas, maka pada penelitian ini diperlukan suatu aplikasi dengan tujuan untuk menentukan klasifikasi kualitas retensi surfaktan dengan metode yang di gunakan adalah k-NN.

1.2. Identifikasi Masalah

Berdasarkan latar belakang tersebut, maka dapat diidentifikasi masalah adalah sebagai berikut :

1. Belum ada dilakukan klasifikasi untuk mempermudah pengguna dalam menentukan kualitas retensi surfaktan pada batuan *sandstone* dengan menggunakan metode *k-Nearest Neighbor*.
2. Untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* dibutuhkan waktu sekitar 3-4 hari menggunakan tes *core flood*.

1.3. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah bagaimana membangun sebuah sistem klasifikasi untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* dengan mudah dan cepat.

1.4. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini hanya membuat untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* menggunakan metode K-Nearest Neighbor.
2. Sistem ini menggunakan beberapa kriteria penginputan untuk parameter atau atribut yang mempengaruhi kualitas retensi surfaktan yaitu sebagai berikut :
 - a. *Absolute Permeability*
 - b. *Total Acid Number*
 - c. *Temperature*
 - d. *Co Solvent*
 - e. *Polimer Salinity*
 - f. *Max Effluent pH*
 - g. *Mobility Ratio*
 - h. MW Surfaktan
3. Sedangkan untuk kriteria *output* adalah nilai kualitas retensi surfaktan pada batuan *sandstone* menggunakan metode K-Nearest Neighbor.
4. Yang digunakan adalah data penelitian dari teknik perminyakan sebanyak 100 data.

5. Penelitian ini menggunakan 10 *fold* dan menentukan nilai $k = 3, 5, 7, 9, 11, 13, 15, 17, 19$ dengan metode *k-Fold Cross Validation* untuk menentukan tingkat akurasi.

1.5. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Melakukan klasifikasi untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* menggunakan metode *K-Nearest Neighbor*.
2. Membuat sebuah sistem aplikasi untuk menentukan kualitas retensi surfaktan berdasarkan hasil klasifikasi yang bersifat *user friendly*.

1.6. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Membantu dan mempermudah bagi pengguna (mahasiswa, dosen, tenaga ahli, dll) untuk menentukan kualitas retensi surfaktan pada batuan *sandstone*.
2. Membantu pengguna (mahasiswa, dosen, tenaga ahli, dll) dalam memperkecil kesalahan data pada proses *input* data, proses dan *output* hasil.

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam penelitian ini, penulis mengambil beberapa referensi dari penelitian-penelitian yang ada sebelumnya. Adapun penelitian sebelumnya adalah sebagai berikut :

Penelitian yang dilakukan oleh Fadhila Tangguh Admojoa, Ahsanawati (2020) membuat sebuah sistem klasifikasi aroma alkohol menggunakan metode KNN. Hasil dari penelitian ini keseluruhan data berhasil diklasifikasikan dengan baik sehingga memperoleh rata-rata 95.8% untuk $k=3$ dan 96.4% untuk $k=4$. Persamaan penelitian ini dengan penelitian diatas adalah sama-sama menggunakan metode k-NN dan menggunakan *k-Fold Cross Validation* untuk menentukan tingkat akurasi. Sedangkan perbedaan penelitiannya terletak pada kasusnya, pada penelitian diatas membahas tentang klasifikasi aroma alkohol dan untuk penelitian ini membahas tentang klasifikasi kualitas retensi surfaktan.

Dalam penelitian Riyan Latifahul Hasanah, Muhamad Hasan, Witriana Endah Pangesti, Fanny Fatma Wati, Windu Gata (2019) menggunakan metode k-NN yang merupakan sebuah metode untuk melakukan klasifikasi penerima bantuan dana desa. penelitian ini, terdapat 159 record atau tuple dengan empat atribut dan hasilnya menunjukkan bahwa dengan $K=15$ dan $K=30$ data baru (D160) memiliki kategori “Tidak Layak” dengan tingkat akurasi sebesar 100%. Kemudian dengan

K=45, K=60 dan K=75 data baru (D160) memiliki kategori “Layak” dengan tingkat akurasi sebesar 81,25%. Persamaan penelitian ini dengan penelitian tersebut yaitu menggunakan metode k-NN dengan menghitung jarak menggunakan *Euclidean Distance*. Sedangkan perbedaan penelitian tersebut adalah pada penelitian ini membahas tentang implementasi metode k-NN pada retensi surfaktan pada batuan *sandstone* sedangkan penelitian tersebut membahas tentang penerapan metode k-NN untuk penerima bantuan dana desa.

Dalam penelitian Huzain Azis, Purnawansyah, Farniwati Fattah, dan Inggrianti Pratiwi Putri (2020) menjelaskan tentang performa klasifikasi K-NN dan *Cross-validation* pada data pasien pengidap penyakit jantung. Digunakan sebanyak 1000 record terdiri dari 11 atribut (*age, gender, height, dll*). Hasil penelitian dari simulasi rasio dataset 50:50 memperoleh nilai akurasi 82%, presisi 82% , recall 82% dan f-measure 80% pada nilai K=13. Persamaan penelitian huzain aziz dkk diatas dengan penelitian ini adalah sama-sama menggunakan k-NN dan untuk tingkat akurasi menggunakan *Cross Validation*. Sedangkan perbedaan penelitian diatas ini terletak pada kasusnya, jika penelitian diatas membahas tentang performa klasifikasi k-NN dan *Cross Validation* pada data pasien pengidap penyakit jantung sedangkan pada penelitian ini membahas tentang data perminyakan mengenai klasifikasi kualitas retensi surfaktan pada batuan *sandstone*.

Dalam Penelitian yang dilakukan oleh Dewi Cahyantia, lifah Rahmayania, Syafira Ainy Husniara (2020) menjelaskan penelitian tentang analisis performa metode k-NN pada dataset pasien pengidap kanker payudara. Diperoleh hasil tertinggi untuk Akurasi dengan nilai 0,93 pada 20% keempat (K3), 20% Pertama

(K4) dan 20% pertama (K5), untuk Presisi dengan nilai 0,97 pada 20% keempat (K3), untuk Recall dengan nilai 0,98 pada 20% ketiga (K3) dan F-measure dengan nilai 0,94 pada 20% keempat (K3) dan 20% ketiga (K5).

Persamaan penelitian ini dengan penelitian tersebut adalah menggunakan k-NN dan yang menjadi perbedaan adalah pada penelitian ini menambahkan metode *k-Fold cross validation* pada uji akurasi datanya.

2.2 Dasar Teori

2.2.1 Enhanced Oil Recovery (EOR)

Menurut Roger & Lee (2012) banyak perusahaan baik produsen internasional maupun nasional, dihadapkan pada tantangan peningkatan cadangan dan produksi minyak. Tantangan untuk EOR biasanya adalah salah satu pengelolaan risiko, sehingga investasi dapat dilakukan dengan diterima kepastian mencapai tujuan yang diinginkan. Metode EOR ini masih baru, sering tidak dilihat di bidang investor. Sejarah kinerja kurang mendukung dan sistemnya lemah. Kebutuhan pelatihan untuk insinyur, operator dan manajer tinggi.

Tujuan dilakukannya EOR adalah:

1. Meningkatkan faktor perolehan minyak.
2. Mengurangi saturasi minyak residual.
3. Menurunkan viskositas minyak yang terdapat dalam reservoir.
4. Memberikan *driving force* pada laju produksi minyak yang sudah rendah.
5. Meningkatkan areal *sweep effecinecy*.

Menurut Abdelgawad & Mahmoud (2014) dengan meningkatnya aktivitas di industri minyak, semakin banyak perhatian diberikan pada pemulihan minyak yang

ditingkatkan *Enhanced Oil Recovery* (EOR). Peningkatan pemulihan minyak EOR disebut metode pemulihan minyak tersier. Dimana menggabungkan metode yang digunakan untuk meningkatkan pemulihan minyak akhir dengan menginjeksikan fluida yang sesuai yang biasanya tidak ada di *reservoir*, seperti bahan kimia, pelarut, oksidasi dan panas untuk menginduksi mekanisme baru untuk memproduksi sisa minyak yang merupakan target dari proses EOR.

2.2.2 *Chemical Flooding*

Chemical Flooding adalah suatu metode yang dapat meningkatkan produksi minyak bumi tersebut. *Chemical Flooding* adalah suatu proses penginjeksian bahan kimia seperti; surfaktan, *polymer*, atau *alkaline* di *reservoir* untuk mengeluarkan *residual oil* dengan cara meningkatkan mobilitas, menurunkan *interfacial tension* dan meningkatkan efisiensi penyapuan (Ahmadi, 2014; Park, Lee, Rosli, & Sulaiman, 2015). *Chemical Flooding* adalah salah satu metode EOR terbaik yang mampu memulihkan 35% dari OOIP (*Original Oil in Place*) (Mandal, 2015). *Chemical Flooding* terdiri dari beberapa metode, salah satunya, yaitu: *Surfactant Flooding*. Surfaktan merupakan komponen organik yang amphiphilic. Rantai hidrokarbon terdiri dari kepala dan ekor (*hydrophilic* merupakan bagian kepala, *hydrophobic* merupakan bagian ekor). Oleh karena itu, surfaktan dapat larut pada solven dan air. Penyerapan surfaktan pada permukaan atau antar muka fluida akan mengubah sifat permukaan yang signifikan sehingga dapat mengurangi tegangan permukaan atau *interfacial tension* (IFT) (Sheng, 2011).

2.2.3 Surfaktan

Surfaktan atau *surface active agent* adalah molekul-molekul yang mengandung gugus hidrofilik (suka air) dan lipofilik (suka minyak/lemak) pada molekul yang sama (Sheat dan Foster, 1997).

Penggunaan surfaktan sebagai metode *oil recovery* sudah diimplementasikan lebih dari 35 tahun di *reservoir*, khususnya di *United States of America* (Iglauer, Wu, Shuler, Tang, & Iii, 2010; Kamari et al., 2015). Injeksi surfaktan ini ditujukan untuk memproduksi *residual oil* yang ditinggalkan oleh *water drive*, dimana minyak yang terjebak oleh tekanan kapiler tersebut tidak dapat bergerak. Minyak tersebut dapat dikeluarkan dengan menginjeksikan larutan surfaktan. Surfaktan bekerja dengan cara menurunkan *interfacial tension* antara minyak dan batuan, hingga minyak yang terperangkap tersebut dapat terlepas dari batuan. Selain itu, surfaktan juga diharapkan dapat merubah *wettability* batuan, dari yang *oil-wet* menjadi *water-wet*, sehingga minyak tersebut dapat lebih mudah untuk terlepas dari batuan. Setelah minyak dapat bergerak, maka diharapkan tidak ada lagi minyak yang tertinggal. *Surfactant flooding* adalah *chemical flooding* yang paling menjanjikan, namun injeksi ini bersifat tidak ekonomis. Hal tersebut karena surfaktan mudah terserap pada permukaan (matriks) batuan (Park et al., 2015; Thomas, 2007; Wang, 1993). Kejadian tersebut dinamakan retensi surfaktan. Retensi surfaktan adalah permasalahan utama pada saat penginjeksian surfaktan, yang mana penginjeksian tersebut berfungsi untuk memulihkan *residual oil* (Glover, Puerto, Maerker, & Sandvik, 1979). Hilangnya surfaktan tersebut dapat berdampak pada hilangnya nilai guna dari surfaktan dan hilangnya nilai ekonomis

yang diharapkan (Chengzhi, Bazin, Labrid, & Yanli, 1988; Kamari et al., 2015; Yassin, Arabloo, Shokrollahi, & Mohammad, 2013). Ketika surfaktan tersebut sudah terserap dan tertinggal di matriks batuan, surfaktan tersebut sudah tidak dapat digunakan, sehingga surfaktan tersebut tidak dapat bekerja secara optimal untuk menurunkan *interfacial tension*. Hal ini akan menyebabkan kerugian karena mengingat harga surfaktan yang tidak murah.

Permasalahan retensi surfaktan sudah lama dikenal oleh dunia (Novosad, 1982). Peneliti terdahulu meneliti mengenai retensi surfaktan sebatas parameter yang mempengaruhi surfaktan ataupun menghitung bagaimana cara menghitung retensi surfaktan tersebut. Contohnya pada tahun 2015, Kamari membahas apakah parameter yang paling mempengaruhi nilai retensi surfaktan.

Kamari (2015) berpendapat bahwa nilai dari retensi surfaktan dipengaruhi oleh beberapa parameter, yaitu:

1. ***Permeability Absolute (Kabs)***

Permeability Absolute atau yang disebut juga dengan Kabs adalah kemampuan batuan reservoir untuk dapat meloloskan fluida reservoir melalui pori batuan yang saling berhubungan tanpa merusak partikel pembentuk batuan tersebut. jadi permeabilitas merupakan tingkat kemudahan mengalirnya fluida melalui pori-pori batuan. pada umumnya hasil analisa sample core diperoleh dari reservoir memberikan harga permeabilitas yang berbeda. hal ini menunjukkan sifat ketidakseragaman dari batuan reservoir tersebut. Adapun satuan *Absolute Permeability* adalah mD.

2. *Total Acid Number*

Total Acid Number (TAN) adalah ukuran keasaman yang ditentukan oleh jumlah kalium hidroksida dalam miligram yang diperlukan untuk menetralkan asam dalam satu gram minyak.

3. *Temperature*

Temperature adalah suatu ukuran dingin atau panasnya keadaan atau sesuatu lainnya. Satuan ukur dari *temperature* yang banyak digunakan di Indonesia adalah $^{\circ}\text{C}$ (derajat Celsius). Sementara satuan ukur yang banyak digunakan diluar negeri adalah derajat Fahrenheit (Ir, Sarsinta,2008). Pada penelitian ini *temperature* yang dimaksud adalah suatu ukuran tingkat derajat panas dalam *reservoir*. Reservoir adalah tempat terakumulasinya minyak dan gas didalam bumi, yang dapat berbentuk perangkap struktur (structural trap) atau perangkap stratigrafi (stratigraphical trap).

Untuk formulasi yang diberikan pada kondisi tertentu, peningkatan suhu mengurangi atau meningkatkan retensi surfaktan tergantung pada perilaku fase mikroemulsi dan stabilitas air. Novosad (1982) mengamati penurunan retensi surfaktan dari 1,7 mg/g-batuan menjadi 0,65 mg/g-batuan untuk perubahan suhu dari 25 $^{\circ}\text{C}$ menjadi 70 $^{\circ}\text{C}$.

4. *Co-solvent*

Co-solvent atau *Concentration solvent* adalah zat yang untuk melarutkan zat lain untuk membentuk suatu larutan. Dimana larutan adalah campuran homogen dalam tingkat molekul, apabila kedua zat dicampur pada suatu perbandingan, maka zat yang lebih banyak disebut pelarut dan zat lainnya disebut zat pelarut. Pelarut

dapat berupa gas atau zat padat, namun yang paling sering digunakan adalah cairan. Contohnya air, alkohol dan lainnya. Pelarut dapat berperan penting dalam industri maupun laboratorium.

Berapa sifat-sifat penting lainnya pada pelarut antara lain:

1. Kemampuan melarutkan (*solubility*) yaitu pelarut akan lebih dapat melarutkan zat terlarutnya.
2. Volatility atau kemudahan menguap. Kecepatan menguap pelarut biasanya lebih mudah menguap dari zat terlarutnya.
3. Pelarut biasanya memiliki titik didid yang lebih rendah dari zat terlarutnya.
5. ***Polimer Sanility***

Polimer atau sering disebut makromolekul adalah rangkaian molekul sederhana berukuran sangat panjang yang terbentuk dari perulangan unit-unit kimia kecil dan sederhana (Chauveteau, 1991). Polyacrylamide merupakan polimer sintesis yang bersifat non-ionik yang disintesis dari monomer acrylamide, dan umumnya dibuat berdasarkan mekanisme radikal bebas. Molekul polyacrylamide adalah molekul yang sangat fleksibel, dimana rantai yang panjang dan diameter molekul yang relatif kecil membuat polimer ini sensitif terhadap kerusakan mekanis dan degradasi. Untuk menambah keefektifannya dalam peningkatan perolehan minyak, polyacrylamide direaksikan dengan basa kuat (NaOH dan KOH) sehingga disebut dengan partially hydrolyzed polyacrylamide (PHPA) (Kasimbazi, 2014). Injeksi menggunakan polimer untuk meningkatkan perolehan minyak pertama kali diperkenalkan sebagai teknik Enhanced Oil Recovery pada awal tahun 1960-an.

6. *Max Effluent pH*

Derajat keasaman (pH) merupakan salah satu karakteristik surfaktan. Setiap jenis surfaktan memiliki pH yang berbeda-beda, misalnya saja pH dari surfaktan dietanolamida berkisar antara 9 dan 10. Dalam penggunaan surfaktan, pH perlu diperhatikan karena akan berpengaruh terhadap aktivitas surfaktan tersebut meskipun ada sebagian jenis surfaktan yang tidak dipengaruhi oleh perubahan pH (Herawan, 1998; Warwel et al., 2001).

7. *Mobility Ratio*

Mobility Ratio adalah ratio yang menyatakan kemudahan suatu fluida untuk dapat bergerak. Perbandingan mobilitas merupakan elemen utama dalam mendesain waterflood yang mana merupakan indikator untuk menentukan besarnya efisiensi penyapuan. Menurut hukum Darcy mobilitas merupakan perbandingan antara permeabilitas efektif fluida dengan viskositas fluida. Pada waterflood perbandingan mobilitas didefinisikan sebagai perbandingan antara mobilitas fluida pendesak (air) dengan mobilitas fluida yang didesak (minyak).

Batasan harga mobility ratio terdiri atas:

1. Jika $M = 1$, menunjukkan bahwa minyak dan air bergerak dengan laju yang relatif sama
2. Jika $M < 1$, menunjukkan bahwa air bergerak dengan laju yang lebih lambat daripada minyak sehingga efisiensi pendesakan lebih tinggi dengan saturasi air pada saat breakthrough lebih tinggi
3. Jika $M > 1$, menunjukkan bahwa air bergerak lebih cepat dibandingkan dengan minyak. Hal ini menyebabkan turunnya efek pendesakan minyak oleh

air.

8. *Molecular weight* (MW) Surfaktan

Molecular weight (MW) disebut juga dengan bobot molar dalam surfaktan. *Molecular weight* (MW) Surfaktan adalah massa dari satu mol senyawa surfaktan yang digunakan untuk penginjeksian surfaktan. Kebijakan konvensional adalah bahwa adsorpsi meningkat seiring dengan berat molekul

surfaktan meningkat. Somasundaran dan Hanna (1977) mengamati peningkatan 128 adsorpsi dengan peningkatan panjang rantai karbon alkil amonium asetat dan sulfonat karena peningkatan interaksi lateral antara rantai.

9. *Surfactant Retention*

Surfactant Retention adalah kemampuan surfaktan untuk terserap kedalam batuan (Chengzhi, Bazin, Labrid, & Yanli, 1988; Kamari et al., 2015; Yassin, Arabloo, Shokrollahi, & Mohammad, 2013).

Adapun pengelompokan nilai retensi surfaktan berdasarkan analisis retensi surfaktan Sriram Solairaj (2012) dapat dilihat pada tabel 2.1

Tabel 2.1 Pengelompokan Retensi Surfaktan

Retensi (mg/g-Rock)	Kriteria
0,16-0,40	Not Good
0,01-0,15	Good

2.2.4. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari (Hermawati, 2013).

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu:

1. Fungsi Prediksi (*prediction*).

Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis.

2. Fungsi Deskripsi (*description*).

Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.

3. Fungsi Klasifikasi (*classification*).

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.

4. Fungsi Asosiasi (*association*).

Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

2.2.5. Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Hans & Kamber, 2006). Klasifikasi termasuk dalam kelompok *supervised learning*, yaitu sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011):

- a. Kelas. Variabel yang berupa kategori yang merepresentasikan “label” yang terdapat pada objek.
- b. Predictor. Variabel yang direpresentasikan oleh karakteristik (atribut) data.
- c. *Training* dataset. Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
- d. *Testing* dataset. Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Beberapa metode penilaian pada klasifikasi adalah seperti berikut:

- a. *Leave-one-out*, setiap data digunakan untuk pengujian sekali pada sebanyak model yang dikembangkan sehingga ada sejumlah titik data. Metode

memakan waktu yang lama, namun untuk dataset yang kecil metode ini adalah pilihan yang tepat.

- b. *Bootstrapping*, sejumlah instans yang tetap dari data awal dijadikan sampel (dengan pergantian) untuk pelatihan dan dataset sisanya digunakan untuk pengujian. Proses ini diulang-ulang sebanyak yang diinginkan.
- c. *Jackknifing*, akurasi dihitung dengan mengeluarkan satu sampel pada setiap iterasi dari proses estimasi.
- d. *Area under the ROC curve*, adalah teknik penilaian secara grafis dimana true positive rate digambar pada sumbu Y dan false positive rate digambar pada sumbu X.
- e. *k-fold cross validation*, dataset yang utuh di pecah secara acak menjadi beberapa kelompok dengan *size* yang hampir sama dan semua data pada kelompok tersebut dilatih kecuali data pada kelompok pengujian.

Berikut ini adalah algoritma klasifikasi data mining yang paling populer menurut Gorunescu (2011):

- a. *Decision Trees*.

Decision tree analysis (atau analisa pohon keputusan adalah suatu teknik yang termasuk keluarga machine-learning) bisa dibilang teknik klasifikasi yang paling populer pada area data mining.

- b. *Bayesian classifiers/Naive Bayes classifiers*.

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa

sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas.

c. *Artificial Neural Networks.*

Jaringan syaraf tiruan merupakan simulasi dari otak biologis yang bertujuan untuk belajar mengenali pola data dan mensimulasikan proses belajar adaptif biologis, walau dalam skala yang sangat sederhana.

d. *Rough sets.*

Metode ini mempertimbangkan keanggotaan parsial dari label-label kelas ke kategori-kategori yang telah ditetapkan sebelumnya dalam membuat model-model (kumpulan aturan) untuk problem klasifikasi.

e. *Rule-Based methods.*

Adalah *advanced computer program* yang mencoba untuk meniru kemampuan manusia dalam membuat keputusan dan pemecahan masalah.

f. *Case Based Reasoning.*

Merupakan suatu metode penalaran komputer yang menggunakan pengetahuan lama untuk mengatasi masalah baru yang ditemui. *Case Based Reasoning* memberikan solusi terhadap kasus baru berdasarkan kedekatan antara kasus lama yang ada pada basis pengetahuan dengan kasus baru.

g. *Support Vector Machines.*

SVM adalah metode machine learning yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. SVM merupakan sistem

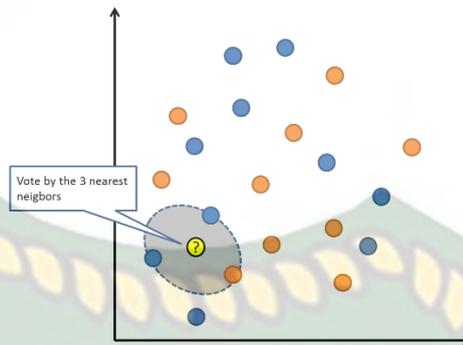
pembelajaran yang menggunakan ruang hipotesis berupa fungsi- fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning* bias yang berasal dari teori pembelajaran statistik.

h. *k-Nearest Neighbor*.

k-Nearest Neighbor (k-NN) adalah suatu metode yang menggunakan algoritma *supervised* untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample*.

2.2.6. Metode *k-Nearest Neighbour*

Algoritma *k-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. k-NN termasuk algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada k-NN. Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample*.



Gambar 2.1 Contoh Klasifikasi pada *k-Nearest neighbor*

Algoritma *k-Nearest Neighbor* menggunakan klasifikasi ketetanggaan (*neighbor*) sebagai nilai prediksi dari query instance yang baru. Algoritma ini sederhana, bekerja berdasarkan jarak terpendek dari query instance ke *training* sample untuk menentukan ketetanggaannya (Rizal, 2013). Berikut beberapa formula perhitungan jarak yang digunakan dalam algoritma k-NN sebagai berikut:

a. *Euclidean Distance*

Jarak *Euclidean* adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi.

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

b. *Hamming Distance*

Jarak *Hamming* adalah cara mencari jarak antar 2 titik yang dihitung dengan panjang vektor biner yang dibentuk oleh dua titik tersebut dalam block kode biner.

c. *Manhattan Distance*

Manhattan Distance atau Taxicab Geometri adalah formula untuk mencari jarak d antar 2 vektor p, q pada ruang dimensi n .

d. *Minkowski Distance*

Minkowski distance adalah formula pengukuran antar 2 titik pada ruang vektor normal .

Keuntungan algoritma *k-Nearest Neighbor* :

1. Sangat mudah diimplementasi dan dipahami.
2. Sangat non-linear.
3. Efektif untuk menghitung data dalam skala kecil.

Kelemahan algoritma *k-Nearest Neighbor* :

1. Memerlukan nilai k yang optimal.
2. Lambat saat proses prediksi.
3. Rentan terhadap perbedaan rentang variabel.

Algoritma klasifikasi dalam *k-Nearest Neighbor* :

1. Menentukan nilai k (jumlah tetangga paling dekat).
2. Menghitung jarak antara data yang akan dievaluasi dengan semua data pelatihan atau data sampel dengan menggunakan rumus *Euclidean Distance*.
3. Mengurutkan hasil jarak step '2' secara *ascending*.
4. Kumpulkan klasifikasi berdasarkan nilai k .
5. Memasangkan kategori atau kelas yang bersesuaian.

6. Mencari jumlah kelas terbanyak dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

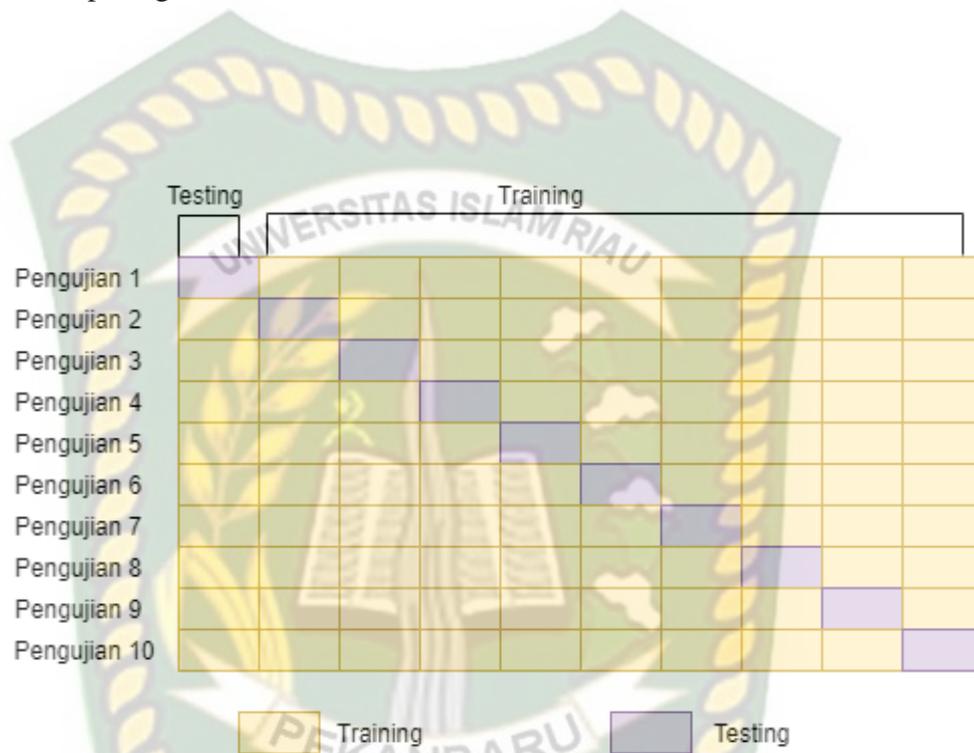
Nilai k yang terbaik untuk algoritma ini tergantung pada data, Salah satu masalah yang dihadapi k -NN adalah dalam pemilihan nilai k yang tepat. Nilai k yang tinggi bisa mengurangi efek noise pada klasifikasi tetapi membuat batasan antara setiap kelas menjadi kabur sedangkan k yang terlalu kecil bisa menyebabkan algoritma terlalu *sensitive* terhadap *noise*. Nilai k terbaik dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *k-fold cross validation*.

2.2.7 . K Fold Cross Validation

Menurut Pandie (2012) *k-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut *input* yang acak. *K-fold cross validation* diawali dengan membagi data sejumlah *fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam n buah partisi dengan ukuran yang sama $D_1, D_2, D_3, \dots, D_n$ selanjutnya proses *testing* dan *training* dilakukan sebanyak n kali. Dalam iterasi ke- i partisi D akan menjadi data *testing* dan sisanya akan menjadi data *training*.

Menurut Hastie (2008), dengan $k=5$ atau $k=10$ dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, sebab data *training* pada setiap *fold* cukup berbeda dengan data *training* yang asli. Secara keseluruhan, 5 atau 10- *fold cross validation* sama-sama direkomendasikan dan disepakati bersama.

Langkah pertama untuk mendapatkan k-optimal dengan pengulangan 10 kali percobaan adalah membagi data *training* sebanyak 10 bagian atau kelompok data yang sama seperti gambar 2.2.



Gambar 2.2 Pembagian data pada metode 10-Fold Cross Validation

Berdasarkan gambar diatas, pada P1 blok pertama terdiri dari data yang akan berperan sebagai data *testing* dan blok lainnya berperan sebagai data *training*. Setiap data *testing* dilakukan prediksi menggunakan algoritma k-NN. Hasil klasifikasi prediksi k-NN dibandingkan dengan klasifikasi data real dan dihitung jumlah prediksi yang cocok dengan data asli. Tingkat akurasi yang tinggi itulah yang terpilih menjadi nilai k terbaik (Banjarsari, 2015).

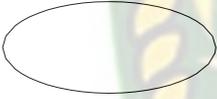
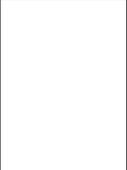
2.3. Tool Dalam Analisa Dan Perancangan Sistem

Dalam membangun sebuah sistem diperlukan beberapa tool dalam menganalisa dan merancang yaitu sebagai berikut ini:

2.3.1 Use Case

Menurut Satzinger, dkk (2012), *use case diagram* adalah diagram yang menggambarkan berbagai peran dari pengguna untuk menggunakan sistem. Tujuan dari penggunaan dari penggunaan *use case diagram* adalah untuk mengidentifikasi bagaimana sistem tersebut akan digunakan. Adapun simbol *use case diagram* dapat dilihat pada tabel 2.42 berikut ini.

Tabel 2.2 Simbol *Use Case Diagram*

Simbol	Nama	Fungsi
	Aktor	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .
	<i>Use Case</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor.
	<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.
<<include>>	<i>Include</i>	Memnspesifikasikan bahwa use case sumber secara eksplisit.
<<extend>>	<i>Extend</i>	Menspesifikasikan bahwa use case target memperluas perilaku dari use case sumber pada suatu titik yang diberikan.
	<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas.

2.3.2 Data Flow Diagram

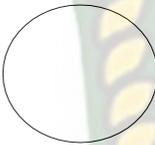
Menurut Rosa A.S. dan M. Shalahuddin (2014), dalam bukunya “Rekayasa Perangkat Lunak”. *Data Flow Diagram* (DFD) adalah representasi grafik yang menggambarkan aliran informasi dan transformasi informasi yang diaplikasikan sebagai data yang mengalir dari masukan (*input*) dan keluaran (*output*).

DFD awalnya dikembangkan oleh Chris Gane dan Trish Sarson pada tahun 1979 yang termasuk dalam *Structured Systems Analysis and Design Methodology* (SSADM) yang ditulis oleh Chris Gane dan Trish Sarson. Sistem yang dikembangkan ini berbasis pada dekomposisi fungsional dari sebuah sistem. Edward Yourdon dan Tom DeMarco memperkenalkan metode yang lain pada tahun 1980an dimana mengubah persegi dengan sudut lengkung (pada DFD Chris Gane dan Trish Sarson) dengan lingkaran untuk menotasikan. DFD Edward Yourdon dan Tom DeMarco populer digunakan sebagai model analisis sistem perangkat lunak untuk sistem perangkat lunak yang akan diimplementasikan dengan pemrograman terstruktur.

DFD sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem yang baru yang akan dikembangkan secara logika dan menjelaskan arus data dari mulai pemasukan sampai dengan keluaran data tingkatan diagram arus data mulai dari diagram konteks yang menjelaskan secara umum suatu sistem atau batasan sistem dari level 0 dikembangkan menjadi level 1 sampai sistem tergambar secara rinci. Gambaran ini tidak tergantung pada perangkat keras, perangkat lunak, struktur data atau organisasi file.

Notasi-notasi pada DFD Edward Yourdon dan Tom DeMarco adalah sebagai berikut :

Tabel 2.3 Notasi-notasi pada DFD Edward Yourdon dan Tom DeMarco

Simbol	Nama	Fungsi
	Simbol entitas eksternal	Digunakan untuk menunjukkan tempat asal <i>data</i> .
	Simbol proses	Digunakan untuk menunjukkan tugas atau proses yang dilakukan baik secara manual atau otomatis
	Simbol penyimpanan <i>Data</i>	Digunakan untuk menunjukkan gudang informasi atau <i>data</i> .
	Simbol arus <i>data</i>	Digunakan untuk menunjukkan arus dari proses.

2.3.3 Context Diagram

Model berikutnya menjawab sejumlah pertanyaan yang muncul dalam pembuatan statement of purpose. *Context Diagram* merupakan kejadian tersendiri dari suatu diagram alir data. Dimana satu lingkaran merepresentasikan seluruh sistem. *Context Diagram* ini harus berupa suatu pandangan, yang mencakup masukan-masukan dasar, sistem-sistem dan keluaran.

Context Diagram merupakan tingkatan tertinggi dalam diagram aliran data dan hanya memuat satu proses, menunjukkan sistem secara keseluruhan. Proses tersebut diberi nomor nol. Semua entitas eksternal yang ditunjukkan pada diagram konteks berikut aliran data-aliran data utama menuju dan dari sistem. Diagram tersebut tidak memuat penyimpanan data dan tampak sederhana untuk diciptakan,

begitu entitas-entitas eksternal serta aliran data-aliran data menuju dan dari sistem diketahui penganalisis dari wawancara dengan user dan sebagai hasil analisis.

Context diagram menggaris bawahi sejumlah karakteristik penting dari suatu sistem:

1. Kelompok pemakai, organisasi, atau sistem lain dimana sistem kita melakukan komunikasi yang disebut juga sebagai terminator.
2. Data dimana sistem kita menerima dari lingkungan dan harus diproses dengan cara tertentu.
3. Data yang dihasilkan sistem kita dan diberikan ke dunia luar.
4. Penyimpanan data yang digunakan secara bersama antara sistem kita dengan terminator. Data ini dibuat oleh sistem dan digunakan oleh lingkungan atau sebaliknya,, dibuat oleh lingkungan dan digunakan oleh sistem kita.
5. Batasan antara sistem kita dan lingkungan.

Context Diagram dimulai dengan penggambaran terminator, aliran data, aliran kontrol penyimpanan, dan proses tunggal yang menunjukkan keseluruhan sistem. Bagian termudah adalah menetapkan proses (yang hanya terdiri dari satu lingkaran) dan diberi nama yang mewakili sistem. Nama dalam hal ini dapat menjelaskan proses atau pekerjaan atau dalam kasus ekstrim berupa nama perusahaan yang dalam hal ini mewakili proses yang dilakukan keseluruhan organisasi.

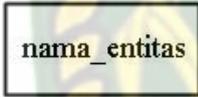
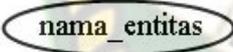
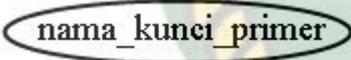
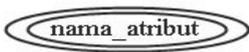
Terminator ditunjukkan dalam bentuk persegi panjang dan berkomunikasi langsung dengan sistem melalui aliran data atau penyimpanan eksternal. Antar terminator tidak diperbolehkan komunikasi langsung.

2.3.4 Entity Relationship Diagram

Entity Relationship Diagram (ERD) adalah sekumpulan cara atau peralatan untuk mendeskripsikan data-data atau objek-objek yang dibuat berdasarkan dan berasal dari dunia nyata yang disebut entitas (*entity*) serta hubungan (*relationship*) antar entitas-entitas tersebut dengan menggunakan beberapa notasi.

Komponen-komponen pembentuk ERD dapat di lihat pada tabel 2.4 di bawah ini.

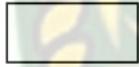
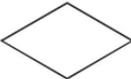
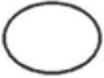
Tabel 2.4 Simbol-Simbol ERD dengan Notasi Chen.

Simbol	Deskripsi
Entitas / <i>entity</i> 	Entitas merupakan data inti yang akan disimpan, bakal tabel pada basis data, benda yang memiliki data dan harus disimpan datanya agar dapat diakses oleh aplikasi komputer, penamaan entitas biasanya lebih ke kata benda dan belum merupakan tabel.
Atribut 	<i>Field</i> atau kolom data yang butuh disimpan dalam suatu entitas.
Atribut kunci primer 	<i>Field</i> atau kolom data yang butuh disimpan dalam suatu entitas dan digunakan sebagai kunci akses <i>record</i> yang diinginkan, biasanya berupa id, kunci primer dapat lebih dari satu kolom asalkan kombinasi dari beberapa kolom tersebut dapat bersifat unik (berbeda tanpa ada yang sama).
Atribut multivalai / <i>multivalue</i> 	<i>Field</i> atau kolom data yang butuh disimpan dalam suatu entitas yang dapat memiliki nilai lebih dari satu.
Relasi 	Relasi yang menghubungkan antar entitas, biasanya diawali dengan kata kerja.
Asosiasi / <i>association</i> 	Penghubung antara relasi dan entitas di mana di kedua ujungnya memiliki <i>multiplicity</i> kemungkinan jumlah pemakaian. Kemungkinan jumlah maksimum terhubung antara entitas satu dengan entitas yang lain disebut dengan kardinalitas. Misalkan ada kardinalitas 1 ke N atau sering disebut dengan <i>one to many</i> menghubungkan entitas A dengan entitas B.

2.3.5 Program Flowchart

Flowchart adalah bagan-bagan yang mempunyai arus yang menggambarkan langkah-langkah penyelesaian suatu masalah. *Flowchart* merupakan cara penyajian dari suatu algoritma (Ladjamudin, 2006). Simbol *Flowchart* dan fungsinya dapat dilihat pada tabel 2.5 berikut.

Tabel 2.5 Simbol dan Fungsi *Flowchart*

No	Simbol	Nama	Fungsi
1		Terminator	Permulaan/pengakhiran program
2		<i>Flow Line</i>	Arah aliran program
3		<i>Preparation</i>	Proses inialisasi/pemberian nilai awal
4		<i>Process</i>	Proses pengolahan data
5		<i>Input/Output Data</i>	Proses <i>input/output</i> data, parameter, informasi
6		<i>Predefined Process</i>	Permulaan sub program / proses menjalankan sub program
7		<i>Decision</i>	Perbandingan pernyataan, penyeleksian data yang memberikan pilihan untuk langkah selanjutnya
8		<i>On Page Connector</i>	Penghubung bagian-bagian <i>Flowchart</i> yang berada pada satu halaman
9		<i>Off Page Connector</i>	Penghubung bagian-bagian <i>Flowchart</i> yang berada pada halaman berbeda

2.3.6 Structured Query Language

Structured Query Language atau *SQL* diawali publikasi makalah E.F Codd (1970) mengenai model relasional "A Relational Model of Data for Large Shared Data Banks" pada tahun 1974. D.Chamberlin dan R.F Boyce mengembangkan bahasa query untuk memanipulasi dan mengekstrak data dari basisdata relasional. Pada tahun 1974, Chamberlin dan Boyce mendefinisikan bahasa "Structured English Query Language" yang diangkat sebagai SEQUEL. dikemukakan dalam makalah berjudul "SEQUEL= A Structured English Query Language". Kemudian SEQUEL berevolusi menjadi versi bernama SEQUEL/2 yang muncul pada tahun 1976. Orang mengejanya menjadi SQL dan menyebutnya dengan "si quel" meski ejaan resmi adalah "s-q-l".

IBM membuat prototipe DBMS *System R* Berdasarkan bahasa SEQUEL/2 untuk memvalidasi kelayakan model relasional. Pada 1980 ,SEQUEL berganti nama menjadi SQL (*Structured Query Language*) . Akar mula yang lain dari bahasa SQL adalah bahasa SQUARE (*Specifying Queries as Relasional Exspression*).SQUARE dirancang sebagai bahasa riset untuk implementasi aljabar relasional dengan kalimat bahasa inggris.

SQL hadir disekitar abad pertengahan tahun 70-an. SQL menerima tambaha-tambahan dan perluasan-perluasan untuk meningkatkan fitur-fitur yang menampung kemampuan DBMS yang dibuat oleh para vendor.

Bahasa SQL mempunyai beberapa bagian, yaitu:

- a. Bahasa pendefinisian data yaitu DDL/ *Data Defenition Languange* untuk pendefinisian skema relasi,penghapusan relasi dan memodifikasi skema relasi

- b. Bahasa manipulasi data interaktif yaitu DML/ *Data Manipulation Language* berisi bahasa *query* berbasis aljabar relasional dan kalkulus relasional, tupel, pemasukan tupel, menghapus tupel, dan memodifikasi tupel.
- c. Pendefinisian view untuk mendefinisikan view.
- d. Kendali transaksi untuk menspesifikasikan permulaan dan akhir transaksi
- e. Integritas, bagian dari DDL untuk menspesifikasikan konstrain-konstrain integritas dimana data disimpan yang harus dipenuhi basisdata. Pembaruan yang melanggar konstrain – konstrain integritas ditolak
- f. Otorisasi, bagian DDL untuk menspesifikasikan hak-hak akses terhadap relasi dan view.

2.4 Pemrograman

Pemrograman adalah proses menulis, menguji dan memperbaiki (*debug*), dan memelihara kode yang membangun sebuah program komputer. Kode ini ditulis dalam berbagai bahasa pemrograman. Tujuan dari pemrograman adalah untuk memuat suatu program yang dapat melakukan suatu perhitungan atau 'pekerjaan' sesuai dengan keinginan si pemrogram (programmer). Adapun sistem ini dibangun dengan bahasa pemrograman PHP.

2.4.1 *Hypertext Preprocessor* (PHP)

Menurut Alexander F.K. Sibero (2011), dalam bukunya “Kitab Suci Web Programming”. *HyperText Proprocessor* (PHP) adalah pemrograman *interpretet* yaitu proses penerjemahan baris kode sumber menjadi kode mesin yang dimengerti komputer secara langsung pada saat baris kode dijalankan. PHP disebut sebagai pemrograman *Server Side Programming*, hal ini dikarena seluruh prosesnya

dijalankan pada server. PHP adalah suatu bahasa dengan hak cipta terbuka atau yang juga lebih dikenal dengan istilah *Open Source*, yaitu penggunaan dapat mengembangkan kode-kode fungsi PHP sesuai dengan kebutuhannya.

Pemrograman PHP dapat ditulis dalam dua bentuk yaitu penulisan baris kode PHP pada *file* tunggal dan penulisan kode PHP pada halaman HTML (embedded). Kedua cara penulisan tersebut tidak memiliki perbedaan, hanya menjadi kebiasaan gaya penulisan dari programmer.

Untuk penulisan kode PHP pada halaman HTML diperlukan tambahan konfigurasi pada web server agar dapat berjalan. Konfigurasi ini bertujuan untuk mendaftarkan ekstensi **.html** agar dapat dikenali *Apache Web Server* dan diproses seperti halnya *file* PHP dengan ekstensi **.php**.

2.4.2 HTML (*Hyper Text Markup Language*)

HTML (*Hyper Text Markup Language*) adalah sekumpulan simbol-simbol atau tag-tag yang dituliskan dalam sebuah file yang dimaksudkan untuk menampilkan halaman pada *web browser*. Tag-tag tadi memberitahu browser bagaimana menampilkan halaman *web* dengan lengkap kepada pengguna. HTML hanya berisi data saja (content), dan untuk mengatur tampilan (layout) halaman menggunakan CSS. Cara kerja HTML sangat sederhana, yaitu berawal dari client yang memanggil berdasarkan URL (*Uniform Resource Locator*) melalui *browser*, kemudian *browser* mendapat alamat dari *web server*, yang nantinya akan memberikan segala informasi yang dibutuhkan web browser. Web browser yang sudah mendapat informasi segera melakukan proses penterjemahan kode HTML dan menampilkannya ke layar pemakai.

2.5 Pengukuran Akurasi

Pengukuran akurasi merupakan tahapan untuk membuktikan tingkat performa suatu algoritma terhadap dataset yang digunakan. Dalam penelitian ini digunakan *confusion matrix* sebagai alat ukur performa algoritma klasifikasi. Tingkat akurasi ini yang nantinya dijadikan acuan para peneliti terkait performa algoritma klasifikasi tersebut. *Confusion Matrix* adalah evaluasi dari sebuah klasifikasi data mining yang direpresentasikan menjadi tabel (Gorunescu 2011). *Convussion matrix* berisi informasi perbandingan label hasil klasifikasi dengan label sebenarnya.

Tabel 2.2 *Confusion Matrix* (Gorunescu 2011)

Classification		Predicted class	
		Class: YES	Class: NO
Observed class	Class YES	a True Positive (TP)	b False Negative (FN)
	Class NO	c False Positive (FP)	d True Negative (TN)

Dari tabel 2.2 dapat dihitung tingkat akurasi dari sebuah model algoritma dengan menggunakan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$\text{Error Rate} = \frac{FP+FN}{TP+TN+FP+FN} \times 100\%$$

Keterangan:

TP Hasil klasifikasi positif dengan kelas sebenarnya positif

FN Hasil klasifikasi negatif dengan kelas sebenarnya positif

FP Hasil klasifikasi positif dengan kelas sebenarnya negatif

TN hasil klasifikasi negatif dengan kelas sebenarnya negatif



BAB III

METODOLOGI PENELITIAN

3.1 Alat dan Bahan Penelitian yang Digunakan

Alat yang digunakan untuk mendukung proses perancangan sistem dalam penelitian ini terdiri dari dua jenis alat, yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*).

1. Perangkat Keras (*Hardware*)
 - a. *Processor* : Intel ® Core™ i3-370M
 - b. RAM : 2.00 GB
 - c. *System type* : 64-bit *Operating System*
2. Perangkat Lunak (*Software*)
 - a. Sistem Operasi : *Microsoft Windows 10*
 - b. Bahasa Pemrograman : PHP
 - c. Desain Logika Program : *Draw.io*
 - d. Basis Data : *XAMPP Control Panel Version 2.5*

3.2 Bahan Penelitian yang Digunakan

Bahan-bahan yang digunakan untuk melengkapi penelitian ini antara lain :

1. Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangkaian penelitian. Adapun metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Wawancara

Wawancara untuk pengumpulan informasi yang melalui tentang penelitian, baik itu kepada Bapak Fiki Hidayat, S.T.,M.Eng selaku dosen teknik perminyakan dan juga kepada mahasiswa bidang perminyakan, ataupun kepada yang mengetahui tentang retensi surfaktan.

b. Studi Pustaka

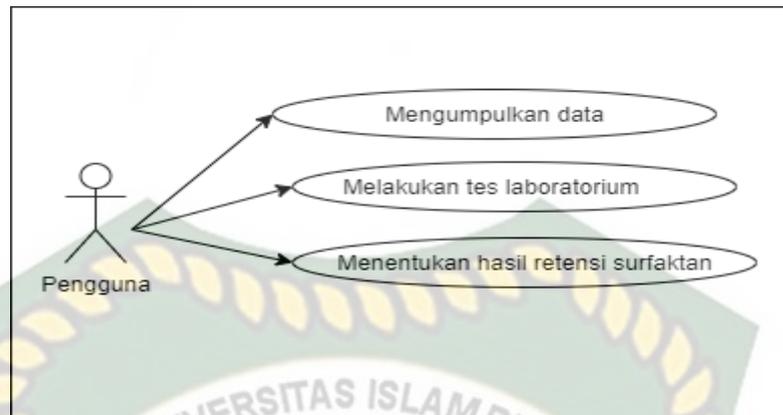
Mengumpulkan data dengan cara mencari dan mempelajari dari berbagai sumber yang berkaitan dengan masalah yang diteliti dalam penyusunan tugas akhir ini, baik dari internet, buku, jurnal ilmiah dan dari bacaan lain yang dapat dipertanggung jawabkan.

2. Jenis Data Penelitian

Adapun jenis data yang digunakan dalam penelitian ini adalah data primer sebanyak 100 data yang didapatkan dari hasil labor teknik perminyakan yang diperoleh langsung dari Bapak Fiki Hidayat, S.T.,M.Eng .

3.3 Analisis Sistem yang Sedang Berjalan

Sebelum melakukan sistem prediksi retensi surfaktan, pengguna melakukan tes laboratorium dan menentukan hasil retensi surfaktan. Analisa sistem yang sedang berjalan bisa dilihat dalam bentuk *Use Case Diagram* pada gambar 3.1.



Gambar 3.1 Use Case Sistem yang Sedang Berjalan

3.4 Perancangan Sistem

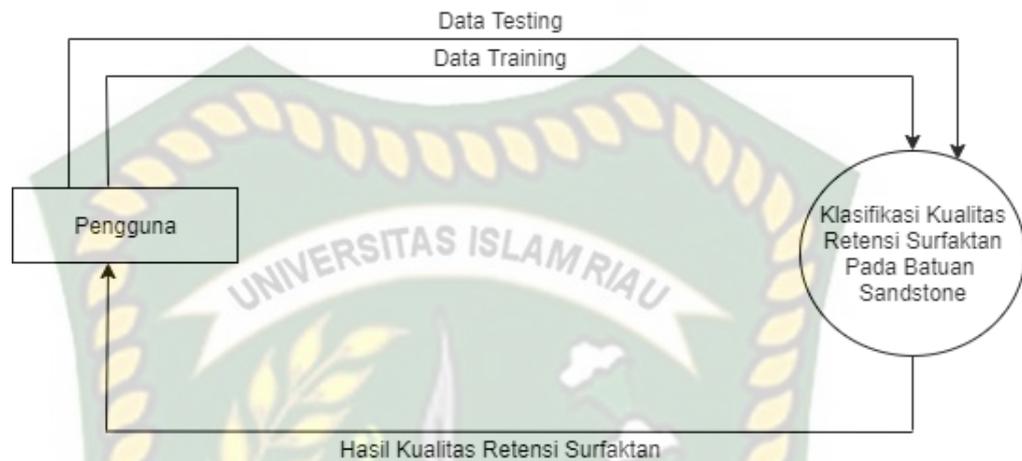
Perancangan sistem ini dibuat sebagai tahapan untuk mempersiapkan proses implementasi sistem yang diinginkan dan untuk menggambarkan secara jelas proses-proses yang diinginkan oleh pengguna. Perancangan sistem informasi bertujuan untuk memberikan gambaran mengenai sistem yang diusulkan sebagai penyempurnaan dari sistem yang sedang berjalan, sistem yang sedang berjalan secara keseluruhan dilakukan secara sederhana sedangkan sistem yang diusulkan akan lebih ditekankan pada pengolahan data secara terkomputerisasi.

Berdasarkan hasil evaluasi terhadap sistem yang berjalan dilihat dari tahapan-tahapan analisis diatas, maka akan dibuat suatu rancangan pengembangan sistem berjalan mengenai klasifikasi kualitas retensi surfaktan pada batuan *sandstone*.

3.4.1 Context Diagram

Context Diagram digunakan untuk menggambarkan hubungan input dan *output* antara sistem dengan entitas luar, suatu diagram konteks selalu memiliki satu proses yang mewakili seluruh sistem. Sistem ini memiliki satu buah eksternal *entity*

yaitu pengguna. Perancangan *Context Diagram* dapat dilihat pada gambar 3.2.

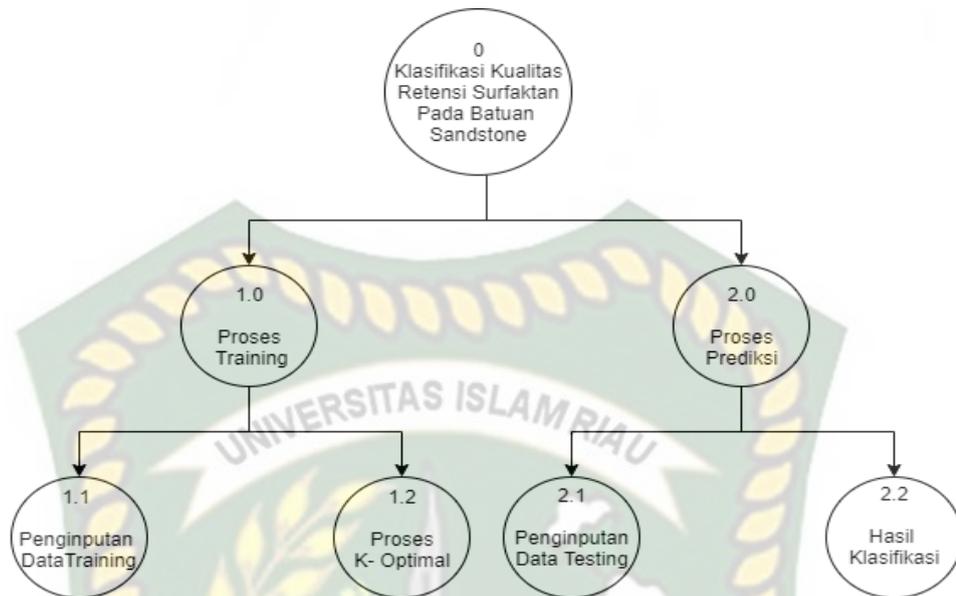


Gambar 3.2 *Context Diagram* Klasifikasi Kualitas Retensi Surfaktan.

Berdasarkan gambar 3.2 seorang pengguna akan menginputkan data *training* dan data *testing* sebagai nilai yang akan diproses oleh sistem dan juga dapat melihat hasil perhitungan dari data tersebut. Data *testing* yang diinput akan diproses oleh sistem dan data tersebut akan menghasilkan sebuah keluaran berupa hasil kualitas retensi surfaktan.

3.4.2 *Hierarchy Chart*

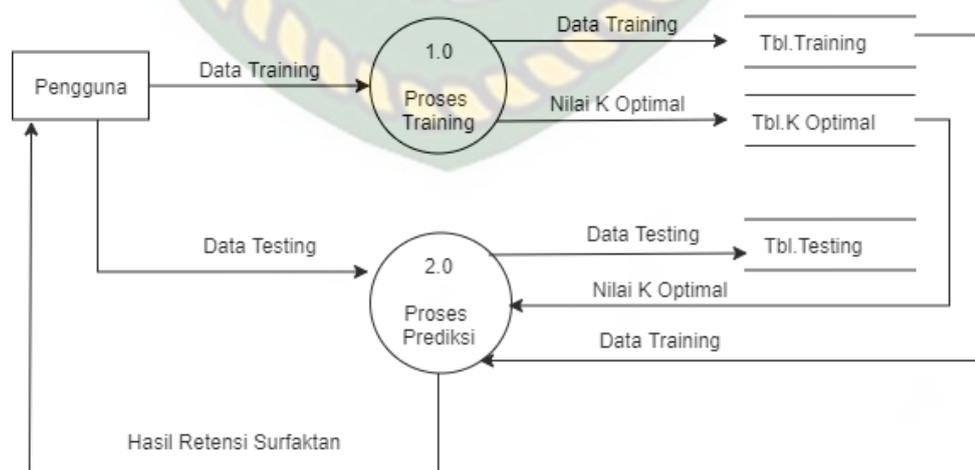
Hierarchy chart digunakan untuk memperlihatkan jenjang atau hirarki dari program yang akan dikembangkan. Dengan demikian dapat dijabarkan urutan kerja dari tiap program. *Hierarchy chart* sistem yang akan dibangun bisa dilihat pada gambar 3.3.



Gambar 3.3 Hierarchy Chart Klasifikasi Kualitas Retensi Surfaktan

3.4.3 Data Flow Diagram (DFD) Level 0

Data Flow Diagram (DFD) berfungsi untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa memperhatikan lingkungan fisik dimana data tersebut mengalir. Rincian dari proses dapat akan diuraikan pada DFD Level 0 seperti pada gambar 3.4.

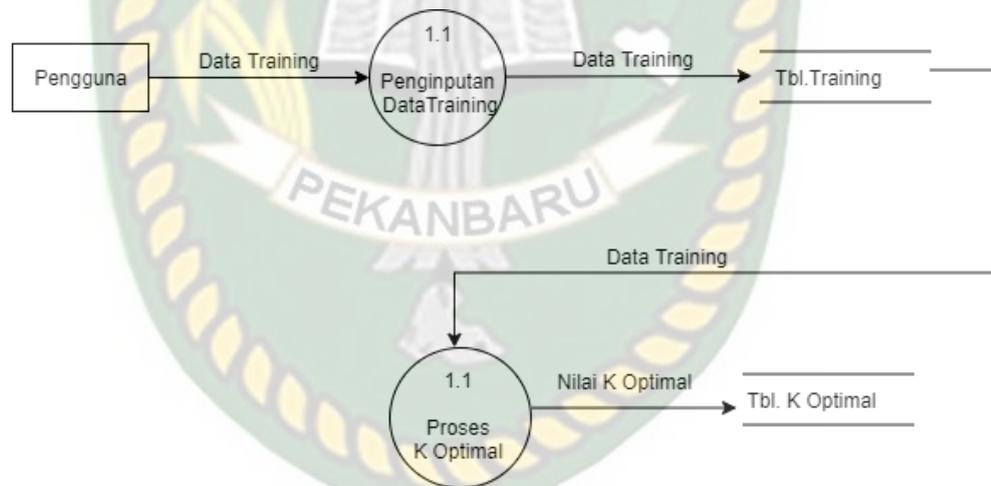


Gambar 3.4 Data Flow Diagram (DFD) Level 0

Berdasarkan gambar 3.4 *DFD* level 0 dapat dilihat aliran data pada sistem yang akan dibangun. Proses pertama yang dilakukan adalah proses *training*. Adapun data yang akan diproses yaitu data *training* dan melakukan proses *k* optimal untuk mendapatkan nilai *k* yang optimal.

3.4.4 Data Flow Diagram (DFD) Level 1 Proses 1

Data Flow Diagram (DFD) level 1 digunakan untuk menggambarkan modul-modul yang ada dalam sistem yang akan dikembangkan. DFD level 1 merupakan *breakdown* DFD level 0 yang sebelumnya sudah dibuat. Rincian dari proses dapat akan diuraikan pada DFD Level 1 proses 1 seperti pada gambar 3.5.

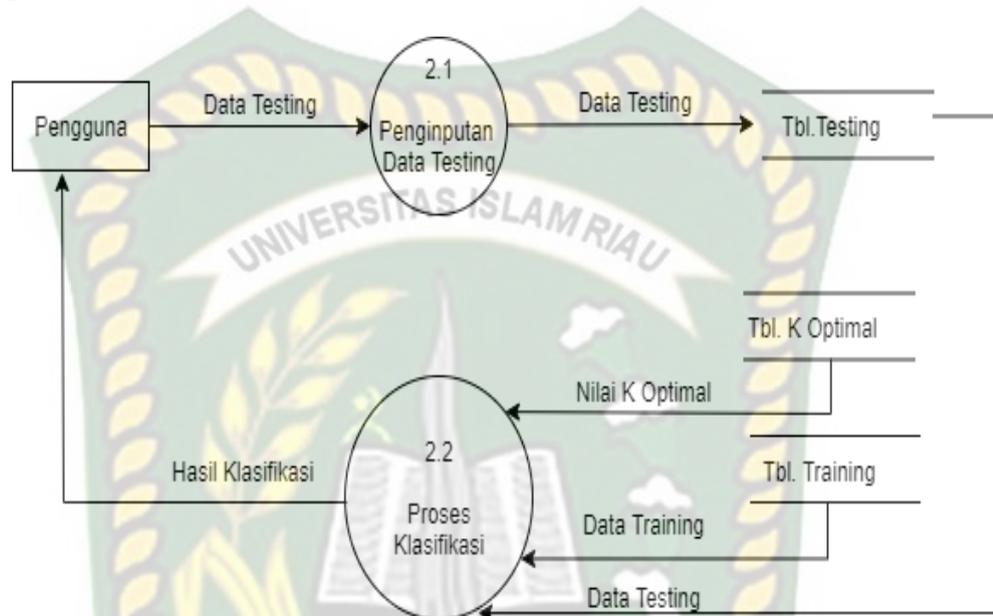


Gambar 3.5 Data Flow Diagram (DFD) Level 1 Proses 1

Berdasarkan gambar 3.5 Data flow diagram level 1 proses 1 menggambarkan proses pengelolaan data *training*. Pada proses 1.1 admin menginput data *training* ke dalam sistem. Selanjutnya akan dibentuk *fold* pada data *training*. Kemudian akan dilakukan proses knn pada setiap *fold* untuk mendapatkan nilai *k* optimal.

3.4.5 Data Flow Diagram (DFD) Level 1 Proses 2

Rincian dari proses dapat akan diuraikan pada DFD Level 1 proses 2 seperti pada gambar 3.6.



Gambar 3.6 Data Flow Diagram (DFD) Level 1 Proses 2

Berdasarkan gambar 3.6, pada proses 2.1 admin menginputkan data *testing* kedalam sistem. Kemudian pada proses 2.2 dilakukan proses knn pada data *testing* yang diinputkan.

3.5 Rancangan Desain

3.5.1 Desain Output

Beberapa desain *output* yang dirancang dalam sistem yang dibangun :

1. Desain Output Data Training

Gambaran *output* untuk melihat data *training* yang sudah diinputkan seperti pada gambar 3.7.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone										Username
Data Training										
No	Kabs	TAN	Temperature	Co Solvent	Polimer Salinity	Max Effluent pH	Mobility Ratio	MW Surfactant	Surfactant Retention	
000	000	000	000	000	000	000	000	000	X(15)	

Gambar 3.7 Desain *Output Data Training*

2. Desain *Output Data Testing*

Gambaran *output* untuk melihat data *testing* yang sudah diinputkan seperti pada gambar 3.8.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone										Username
Data Testing										
No	Kabs	TAN	Temperature	Co Solvent	Polimer Salinity	Max Effluent pH	Mobility Ratio	MW Surfactant		
000	000	000	000	000	000	000	000	000		

Gambar 3.8 Desain *Output Data Testing*

3. Desain *Output* Hasil Perhitungan k-NN

Gambaran *output* untuk melihat hasil perhitungan k-NN dapat dilihat seperti pada gambar 3.9.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone								Username
Hasil Perhitungan KNN								
No	Kabs	TAN	Temperature	Co Solvent	Polimer Salinity	Max Effluent pH	Mobility Ratio	MW Surfactant
000	000	000	000	000	000	000	000	000
Euclidean Distance			Nilai Tertinggi		Kategori			
000			000		X(30)			
Kesimpulan: X (100)								

Gambar 3.9 Desain *Output* Perhitungan k-NN

4. Desain *Output* Nilai K Optimal dan Akurasi

Gambaran *output* untuk nilai k optimal dan akurasi dapat dilihat seperti pada gambar 3.10 berikut.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone			Username																
<p>Nilai K Optimal</p> <table border="1"> <thead> <tr> <th colspan="4">Nilai k (n)</th> </tr> <tr> <th>Pengujian ke-n</th> <th>Data Testing</th> <th>Data Training</th> <th>Kualitas retensi (%)</th> </tr> </thead> <tbody> <tr> <td>X(15)</td> <td>X(20)</td> <td>X(20)</td> <td>999</td> </tr> <tr> <td colspan="4">Rata-rata : 999</td> </tr> </tbody> </table>				Nilai k (n)				Pengujian ke-n	Data Testing	Data Training	Kualitas retensi (%)	X(15)	X(20)	X(20)	999	Rata-rata : 999			
Nilai k (n)																			
Pengujian ke-n	Data Testing	Data Training	Kualitas retensi (%)																
X(15)	X(20)	X(20)	999																
Rata-rata : 999																			
<p>Nilai Akurasi</p> <table border="1"> <thead> <tr> <th>Nilai k</th> <th>Kualitas retensi (%)</th> </tr> </thead> <tbody> <tr> <td>X(15)</td> <td>999</td> </tr> </tbody> </table>				Nilai k	Kualitas retensi (%)	X(15)	999												
Nilai k	Kualitas retensi (%)																		
X(15)	999																		

Gambar 3.10 Desain *Output* Nilai K Optimal dan Akurasi

3.5.2 Desain *Input*

Desain *input* adalah bentuk masukan pada sebuah sistem yang akan diproses untuk menghasilkan sebuah informasi.

1. Desain *Input Login*

Desain input login adalah rancangan yang dibuat untuk peneliti melakukan login terhadap sistem. Rancangan desain login dapat dilihat pada gambar 3.11.

**Klasifikasi Kualitas
Retensi Surfaktan Pada Batuan Sandstone**

Username

Password

Belum Punya Akun ? [Daftar disini](#)

Gambar 3.11 Desain *Input Login*

2. Desain *Input Data Training*

Desain input data *training* merupakan form yang dirancang untuk melakukan input data *training*. Desain input dapat dilihat pada gambar 3.12.

**Klasifikasi Kualitas
Retensi Surfaktan Pada Batuan Sandstone**

Username

Data Training

No	<input type="text" value="999"/>
Kabs	<input type="text" value="999"/>
TAN	<input type="text" value="999"/>
Temperature	<input type="text" value="999"/>
Co Solvent	<input type="text" value="999"/>
Polimer Salinity	<input type="text" value="999"/>
Max Effluent pH	<input type="text" value="999"/>
Mobility Ratio	<input type="text" value="999"/>
MW Surfactant	<input type="text" value="999"/>
Surfactan Retention	<input type="text" value="X (15)"/> ▼

Gambar 3.12 Desain *Input Data Training*

3. Desain *Input Data Testing*

Desain *input data testing* merupakan form yang dirancang untuk melakukan *input data testing*, desain dapat dilihat pada gambar 3.13.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone		Username
Data Testing		
No	<input type="text" value="999"/>	
Kabs	<input type="text" value="999"/>	
TAN	<input type="text" value="999"/>	
Temperature	<input type="text" value="999"/>	
Co Solvent	<input type="text" value="999"/>	
Polimer Salinity	<input type="text" value="999"/>	
Max Effluent pH	<input type="text" value="999"/>	
Mobility Ratio	<input type="text" value="999"/>	
MW Surfactant	<input type="text" value="999"/>	
<input type="button" value="Reset"/> <input type="button" value="Simpan"/>		

Gambar 3.13 Desain *Input Data Testing*

4. Desain *Input Perhitungan K-Fold Cross Validation*

Desain *Input Perhitungan K-Fold Cross Validation* merupakan form yang dirancang untuk melihat nilai k optimal dan akurasi, desain dapat dilihat pada gambar 3.14.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone		Username
<div style="border: 1px solid black; padding: 5px;"> <p>Nilai Akurasi dan K Optimal</p> <hr/> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>Tampilkan Perhitungan Hingga K</p> <input style="width: 100px; text-align: center;" type="text" value="999"/> </div> <div style="text-align: center;"> <input type="button" value="Proses"/> </div> </div> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <input type="button" value="Reset"/> <input type="button" value="Proses"/> </div> </div>		

Gambar 3.14 Desain *Input* Perhitungan *K-Fold Cross Vaidation*

3.5.3 Desain *Database*

Dalam pembuatan sistem ini menggunakan sebuah *database* “ retensi surfaktan” yang terdiri 4 (empat) tabel sebagai berikut:

1. Tabel User

Tabel user merupakan tabel untuk menyimpan data *login*. Tabel ini mempunyai 3 *field*, dimana diantaranya username sebagai *primary key*, *password*, *nama_lengkap* . Rancangan tabel *user* dapat dilihat pada tabel 3.1.

Tabel 3.1 Tabel *User*

No	Nama <i>Field</i>	Tipe Data	Ukuran	Keterangan
1	<i>Username</i>	Varchar	50	<i>Primary key</i>
2	<i>Password</i>	Varchar	50	<i>Password user</i>
3	<i>Nama_Lengkap</i>	Varchar	50	Nama lengkap <i>user</i>

2. Tabel Data *Training*

Tabel *training* merupakan tabel untuk menyimpan data *training*. Tabel ini mempunyai 7 *field*, dimana diantaranya *no_datatrain* sebagai *primary key*, *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio*, *Molecular weight (MW) Surfactant* dan *retensi surfaktan*. Rancangan tabel *training* dapat dilihat pada tabel 3.2.

Tabel 3.2 Tabel *Training*

No	Nama <i>Field</i>	Tipe Data	Ukuran	Keterangan
1	No_datatrain	Int	3	<i>Primary Key</i>
2	Kabs	Float	-	Nilai <i>permeabilitas absolute</i>
3	Tan	Float	-	Nilai <i>total acid number</i>
4	Temp	Float	-	Nilai <i>temperature</i>
5	Solvent	Float	-	Nilai <i>co_solvent</i>
6	Polimer	Float	-	Nilai <i>polimer salinity</i>
7	Eff	Float	-	Nilai <i>max effluent pH</i>
8	Ratio	Float	-	Nilai <i>mobility Ratio</i>
9	Mw	Float	-	Nilai <i>molecular weight surfaktan</i>
10	Retensi	Varchar	15	Kualitas retensi surfaktan

3. Tabel *Testing*

Tabel *testing* merupakan tabel untuk menyimpan data *testing* sebelum disimpan ke tabel *training*. Tabel ini mempunyai 8 *field*, dimana diantaranya *no_datates* sebagai *primary key*, *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant* dan *retensi surfaktan*. Rancangan tabel *training* dapat dilihat pada tabel 3.3.

Tabel 3.3 Tabel *Testing*

No	Nama <i>Field</i>	Tipe Data	Ukuran	Keterangan
1	No_datates	Int	3	<i>Primary Key</i>
2	Kabs	Float	-	Nilai <i>permeabilitas absolute</i>
3	Tan	Float	-	Nilai <i>total acid number</i>
4	Temp	Float	-	Nilai <i>temperature</i>
5	Solvent	Float	-	Nilai <i>co_solvent</i>
6	Polimer	Float	-	Nilai <i>polimer salinity</i>
7	Eff	Float	-	Nilai <i>max effluent pH</i>
8	Ratio	Float	-	Nilai <i>mobility Ratio</i>
9	Mw	Float	-	Nilai <i>molecular weight surfaktan</i>

4. Tabel K Optimal

Tabel k optimal merupakan tabel untuk menyimpan nilai k optimal. Tabel ini mempunyai 4 *field*, yaitu *id_k* sebagai *primary key*, *Distance*, *Rank*, *Akurasi* dan *Tgl*. Rancangan tabel *training* dapat dilihat pada tabel 3.4.

Tabel 3.4 Tabel K Optimal

No	Nama <i>Field</i>	Tipe Data	Ukuran	Keterangan
1	Id_k	Int	3	<i>Primary key</i>
2	Distance	Int	3	Nilai K
3	Rank	Int	3	Nilai tertinggi
4	Akurasi	Float	-	Nilai akurasi
5	Tgl	Date	-	Tanggal

3.6 Perhitungan Manual

A. Perhitungan *K Fold Cross Validation*

Berikut adalah penjelasan perhitungan manual yang terjadi dalam system. Metode *10 fold cross validation* digunakan dalam sistem ini untuk mendapatkan nilai k yang optimal. Pembagian data k-fold adalah sebagai berikut:

Data training : $90 / 10 \text{ fold} = 9$ data per blok.

Dilakukan prediksi menggunakan algoritma k-NN dengan setiap nilai k yang dimasukkan. Dalam penelitian ini, nilai k yang digunakan yaitu 3,5,7,9,11,13,15,17,19 dan hasil perhitungannya dapat dilihat pada tabel 3.5 sampai dengan tabel 3.13.

Tabel 3.5 Tabel Hasil Perhitungan k=3

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	55,56
2	10-18	1-9,19-90	66,67
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	77,78
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	77,78
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			82,22

Tabel 3.6 Tabel Hasil Perhitungan k=5

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	33,33
2	10-18	1-9,19-90	66,67
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	88,89
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	77,78
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata (%)			81,11

Tabel 3.7 Tabel Hasil Perhitungan k=7

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	33,33
2	10-18	1-9,19-90	55,56
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	100
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	88,89
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			77,78

Tabel 3.8 Tabel Hasil Perhitungan k=9

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	33,33
2	10-18	1-9,19-90	44,44
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	77,78
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	88,89
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			74,44

Tabel 3.9 Tabel Hasil Perhitungan k=11

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	44,44	10-90	55,56
2	55,56	1-9,19-90	44,44
3	55,56	1-20,28-90	66,67
4	66,67	1-27,37-90	100
5	55,56	1-36,46-90	77,78
6	100,00	1-45,55-90	100
7	77,78	1-54,64-90	55,57
8	66,67	1-63,73-90	88,89
9	66,67	1-72,82-90	100
10	88,89	1-81	88,89
Rata Rata(%)			77,78

Tabel 3.10 Tabel Hasil Perhitungan k=13

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	22,22
2	10-18	1-9,19-90	55,56
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	77,78
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			75,56

Tabel 3.11 Tabel Hasil Perhitungan k=15

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	0
2	10-18	1-9,19-90	55,56
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	88,89
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			74,44

Tabel 3.12 Tabel Hasil Perhitungan k=17

Pengujian	Data Testing	Data Training	Akurasi Retensi
1	1-9	10-90	11,11
2	10-18	1-9,19-90	55,56
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	88,89
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	77,78
Rata Rata(%)			75,56

Tabel 3.13 Tabel Hasil Perhitungan k=19

Pengujian	Data Testing	Data Training	Akurasi Retensi
-----------	--------------	---------------	-----------------

1	1-9	10-90	0
2	10-18	1-9,19-90	55,56
3	19-27	1-20,28-90	66,67
4	28-36	1-27,37-90	100
5	37-45	1-36,46-90	66,67
6	46-54	1-45,55-90	100
7	55-63	1-54,64-90	55,56
8	64-72	1-63,73-90	100
9	73-81	1-72,82-90	100
10	82-90	1-81	88,89
Rata Rata(%)			73,33

B. Perhitungan *k-Nearest Neighbor*

Berikut proses perhitungan manual yang dilakukan dalam sistem ini. Terdapat data *training* untuk diklasifikasikan dengan data *testing* menggunakan *Absolute Permeability, Total Acid Number, Temperature, Co-solvent, Polimer Salinity, Max Effluent pH, Mobility Ratio* dan *Molecular weight (MW) Surfactant* sehingga dapat menentukan klasifikasi kualitas retensi surfaktan. Adapun data *training* yang digunakan dalam sistem ini dapat dilihat pada tabel 3.5 berikut ini.

Tabel 3.5 Data Training

No	Kabs	Total Number Acid	Temperature	Co Solvent	Polymer Salinity	Max Effluent pH	Mobility Ratio	MW Surfactant	Surfactan Retention
	mD	mg KOH/g-oil	°C	Wt%	ppm	pH	dP wf/dP PD	Mol	
1	207	0	25	0.01	11000	7	0.62	559.51	Tidak Baik
2	195	0	25	0.01	11000	7.5	0.95	559.51	Tidak Baik
3	324	0	25	0.01	15250	6	0.84	559.51	Tidak Baik
4	369	0	25	0.01	8000	7	0.42	559.51	Tidak Baik
..
..
89	304	0.8	85	0	32000	11	0.35	479.84	Baik
90	307	0	30	0.02	20000	7	0.5	559.51	Tidak Baik

Setelah data *training* maka perlu adanya data *testing* dengan atribut *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant* dapat dilihat pada tabel 3.6.

Tabel 3.6 Data Testing

No	Kabs	Total Number Acid	Temperature	Co Solvent	Polymer Sanility	Max Effluent pH	Mobility Ratio	MW Surfactant
	mD	mg KOH/g-oil	°C	Wt%	ppm	pH	dP wf/dP _{PD}	Mol
1	478	0,1	100	0,001	25000	7	0,93	679,91
2	187	0,1	100	0,002	20000	7	0,2	601,6
3	171	0,8	83	0,025	8300	7	0,45	676
4	1470	0,1	46	0,01	5000	11	0,1	368,52
5	1418	0,2	46	0,003	5000	11	0,1	515,19
6	144	0	83	0,2	41000	7	0,29	676,39
7	432	0	55	0	17000	11	0,1	584,67
8	450	0,3	55	0	12600	11	0,08	567,48
9	304	0,5	85	0,0025	4500	11	0,35	479,84
10	300	0,5	55	0,1	32000	10,5	0,1	496,51

1. Penetapan Nilai K

Langkah awal metode *k-Nearest Neighbor* adalah menentukan nilai k. Sebagai contoh k optimal yang digunakan adalah k = 3.

2. Perhitungan Jarak

Untuk Perhitungan jarak pada masing-masing data *training* yang pertama dilakukan adalah mencari nilai terdekat dengan menggunakan rumus Euclidean. Dalam kasus ini maka rumus akan berbentuk seperti ini:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + (x_5 - y_5)^2 + (x_6 - y_6)^2 + (x_7 - y_7)^2 + (x_8 - y_8)^2}$$

Berdasarkan rumus diatas diketahui:

x_1 = *Absolute Permeability* dari data *training*

y_1 = *Absolute Permeability* dari data *testing*

x_2 = *Total Acid Number* dari data *training*

y_2 = *Total Acid Number* dari data *testing*

x_3 = *Temperature* dari data *training*

y_3 = *Temperature* dari data *testing*

x_4 = *Co Solvent* dari data *training*

y_4 = *Co Solvent* dari data *testing*

x_5 = *Polymer Sanility* dari data *training*

y_5 = *Polymer Sanility* dari data *testing*

x_6 = *Max Effluent pH* dari data *training*

y_6 = *Max Effluent pH* dari data *testing*

x_7 = *Mobility Ratio* dari data *training*

y_7 = *Mobility Ratio* dari data *testing*

x_8 = *MW Surfactant* dari data *training*

y_8 = *MW Surfactant* dari data *testing*

Berikut adalah contoh perhitungan jarak dengan menggunakan data *testing* nomor 1.

$$\text{Absolute Permeability} = 478$$

$$\text{Total Acid Number} = 0,1$$

$$\text{Temperatur} = 100$$

$$\text{Co Solvent} = 0,001$$

$$\text{Polymer Sanility} = 25000$$

$$\text{Max Effluent pH} = 7$$

$$\text{Mobility Ratio} = 0,93$$

$$\text{MW Surfactant} = 679,91$$

$$D1 = \sqrt{\begin{aligned} &(207 - 478)^2 + (0 - 0,1)^2 + (25 - 100)^2 + (0,01 - 0,001)^2 + \\ &(11000 - 25000)^2 + (7 - 7)^2 + (0,62 - 0,93)^2 + (559,91 - 679,91)^2 \\ &= 14003,34 \end{aligned}}$$

$$D2 = \sqrt{\begin{aligned} &(195 - 478)^2 + (0 - 0,1)^2 + (25 - 100)^2 + (0,01 - 0,001)^2 + \\ &(11000 - 25000)^2 + (7,5 - 7)^2 + (0,95 - 0,93)^2 + (559,91 - 679,91)^2 \\ &= 14003,58 \end{aligned}}$$

$$D3 = \sqrt{\begin{aligned} &(324 - 478)^2 + (0 - 0,1)^2 + (25 - 100)^2 + (0,01 - 0,001)^2 + \\ &(15250 - 25000)^2 + (6 - 7)^2 + (0,84 - 0,93)^2 + (559,91 - 679,91)^2 \\ &= 9752,248 \end{aligned}}$$

$$D90 = \sqrt{\begin{aligned} &(307 - 478)^2 + (0 - 0,1)^2 + (30 - 100)^2 + (0,02 - 0,001)^2 + \\ &(20000 - 25000)^2 + (7 - 7)^2 + (0,5 - 0,93)^2 + (559,91 - 679,91)^2 \\ &= 5004,861 \end{aligned}}$$

Adapun hasil perhitungan jarak dapat dilihat pada tabel 3.7 berikut ini.

Tabel 3.7 Perhitungan Jarak

No	Kabs mD	Total Number Acid mg KOH/g- oil	Temper ature °C	Co Solvent Wt%	Polymer Sanility ppm	Max Effluent pH pH	Mobility Ratio dP wf/dP PD	MW Surfactant Mol	Surfactan Retention	Jarak
1	207	0	25	0,01	11000	7	0,62	559,51	Tidak Baik	14003,34
2	195	0	25	0,01	11000	7,5	0,95	559,51	Tidak Baik	14003,58
3	324	0	25	0,01	15250	6	0,84	559,51	Tidak Baik	9752,248
4	369	0	25	0,01	8000	7	0,42	559,51	Tidak Baik	17000,94
..
..
89	304	0,8	85	0	32000	11	0,35	479,84	Baik	7005,037
90	307	0	30	0,02	20000	7	0,5	559,51	Tidak Baik	5004,861

3. Pengurutan Data

Setelah perhitungan jarak selesai, maka data tersebut di urutkan berdasarkan jarak terkecil ke besar. Adapun pengurutan data dapat dilihat pada tabel 3.8 dibawah ini.

Tabel 3.8 Pengurutan Data

No	Kabs	Total Number Acid mg KOH/g-oil	Temperature °C	Co Solvent Wt%	Polymer Sanility ppm	Max Effluent pH pH	Mobility Ratio dP wf/dP PD	MW Surfactant Mol	Surfactan Retention	Jarak
1	8	478	0.1	100	0	25000	7	0.93	679.91	0
2	63	478	0.1	100	0	25000	7	0.93	679.91	0
3	42	115	0.5	83	0.001	26000	10	0.07	849.76	1077.46

4 Penentuan Klasifikasi Kualitas Retensi Surfaktan

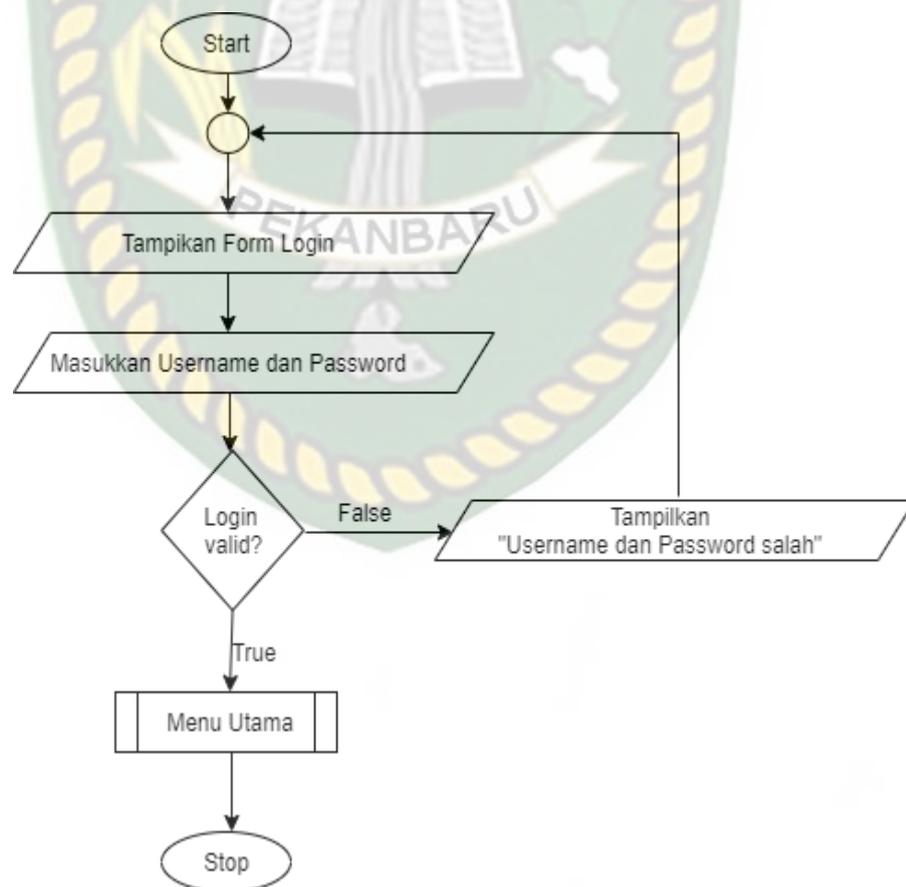
Setelah data diurut, data *filter* sebanyak k optimal yaitu : 3. Maka dari keseluruhan hasil perhitungan jarak dengan data *training* yang berjumlah 90 data, hasil perhitungan diurutkan dari terkecil hingga terbesar. Setelah diurutkan, dilihat mayoritas klasifikasi yang muncul dari perhitungan jarak yang terkecil. Dapat dilihat pada tabel pengurutan data, mayoritas klasifikasi yang muncul adalah “tidak baik”. Sehingga data *testing* yang diklasifikasikan termasuk klasifikasi “tidak baik”.

3.7 Desain Logika Program

Desain logika program menggambarkan logika program yang akan dibuat dengan menggunakan alat bantu program *Flowchart* yang bertujuan untuk menunjukkan alur program, mulai dari data *input* sampai hasil *output* yang diproses oleh program. Adapun logika program dari sistem ini adalah sebagai berikut:

1. Program *Flowchart* Login

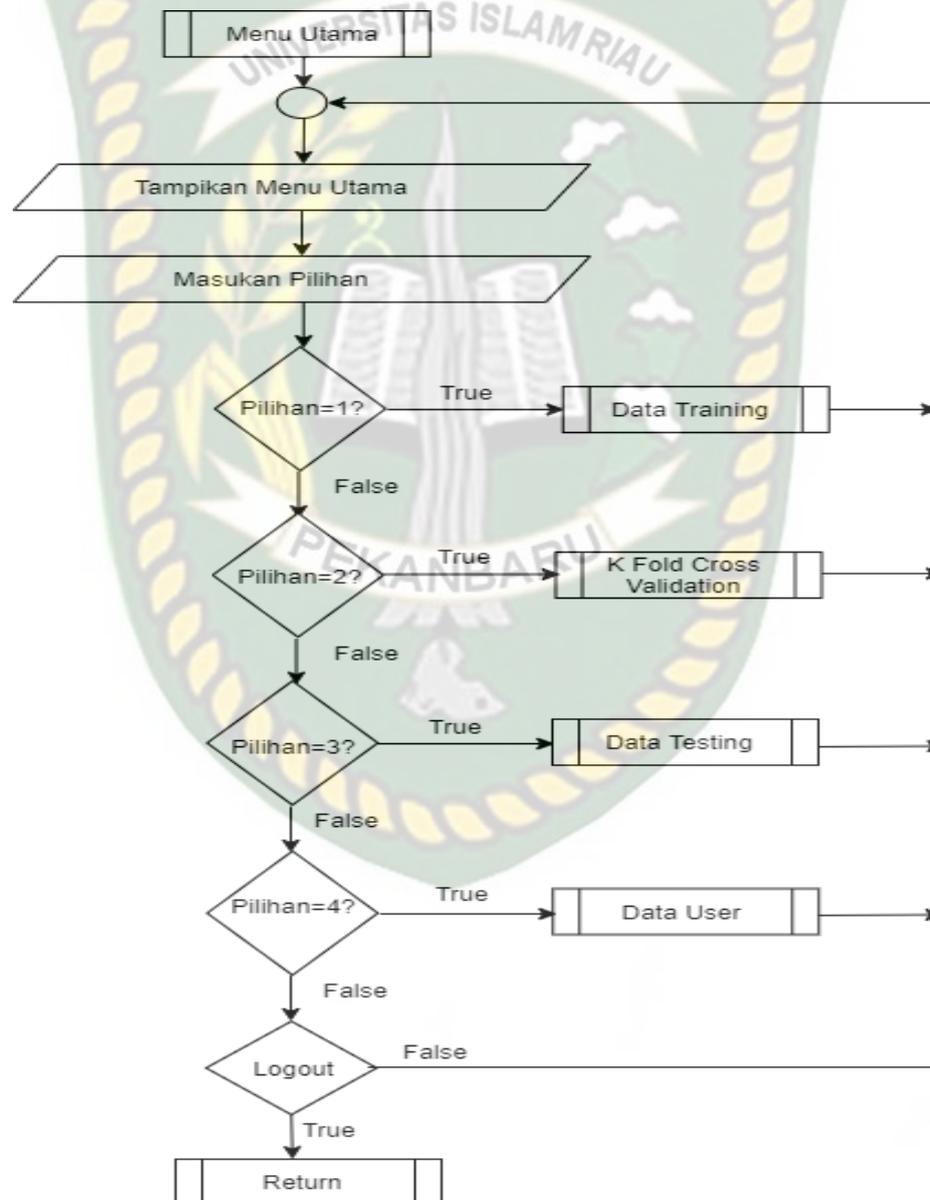
Program *Flowchart* login adalah *Flowchart* yang menjelaskan bagaimana proses admin untuk login ke menu utama sistem. Adapun desain program *Flowchart* login dapat dilihat pada gambar 3.15.



Gambar 3.15 Program *Flowchart* Login

2. Program *Flowchart* Menu Sistem

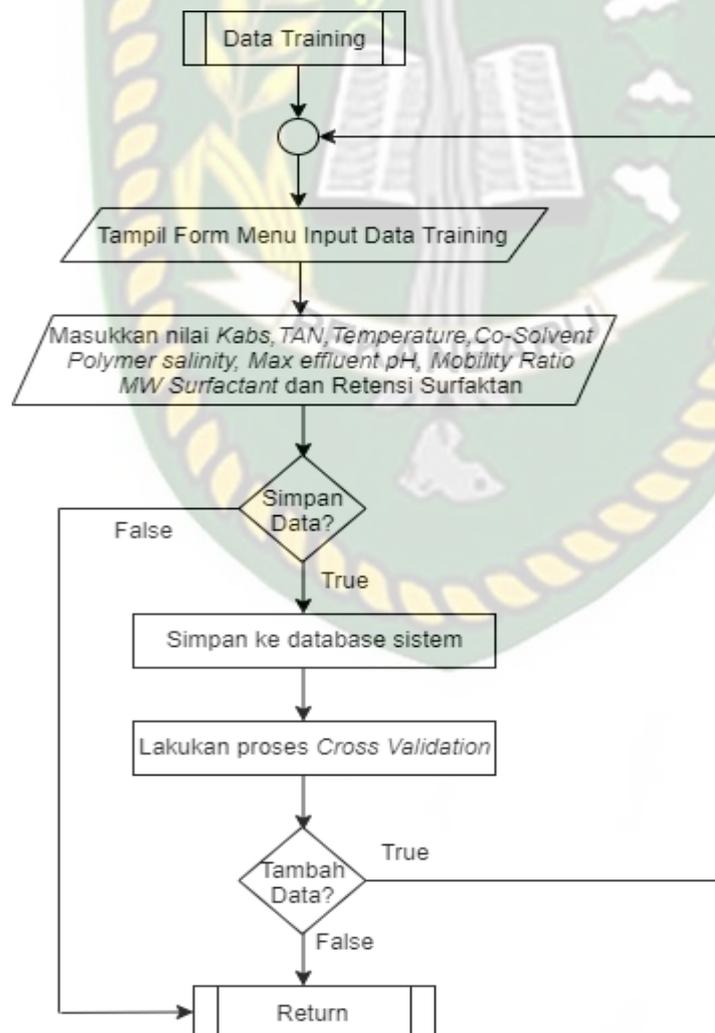
Flowchart ini menjelaskan menu-menu yang terdapat dalam sistem. Menu tersebut terdiri dari menu data *training*, menu data *testing*, menu *k-fold cross validation* dan menu user. Adapun desain program *Flowchart* menu sistem dapat dilihat pada gambar 3.16.



Gambar 3.16 Program *Flowchart* Menu Utama

3. Program *Flowchart* Input Data Training

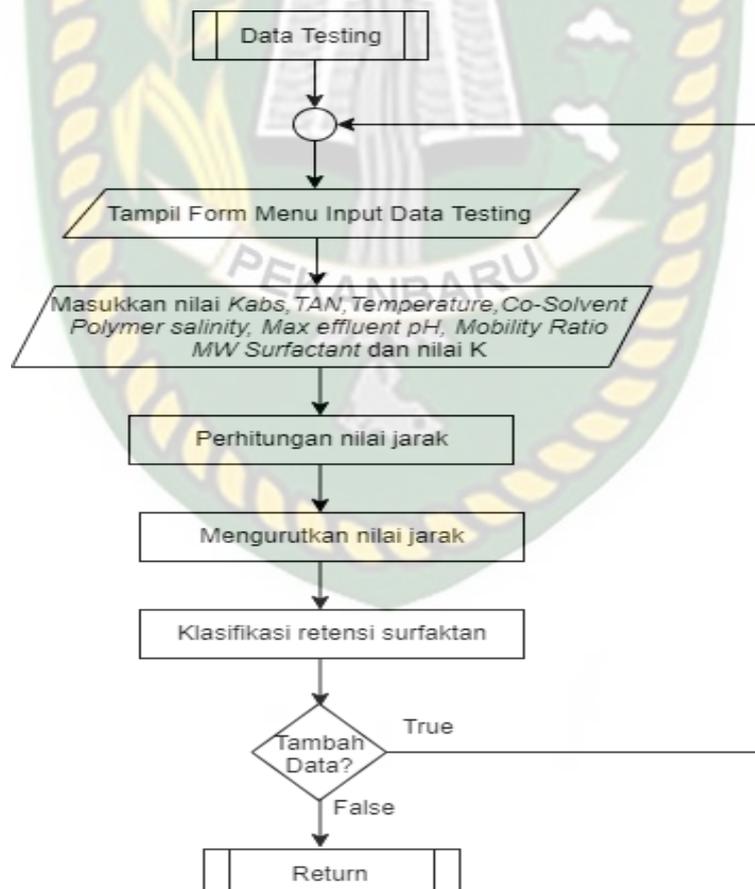
Program *Flowchart* input data training merupakan alur data program yang merancang input data training yang akan tersimpan kedalam sistem. Atribut yang dibutuhkan dalam proses ini *Absolute Permeability, Total Acid Number, Temperature, Co-solvent, Polimer Salinity, Max Effluent pH, Mobility Ratio* dan *Molecular weight (MW) Surfactant* dan Retensi Surfaktan. Adapun desain program *Flowchart* perhitungan data testing dapat dilihat pada gambar 3.17.



Gambar 3.17 Program *Flowchart* Input Data Training

4. Program *Flowchart* Perhitungan Data Testing

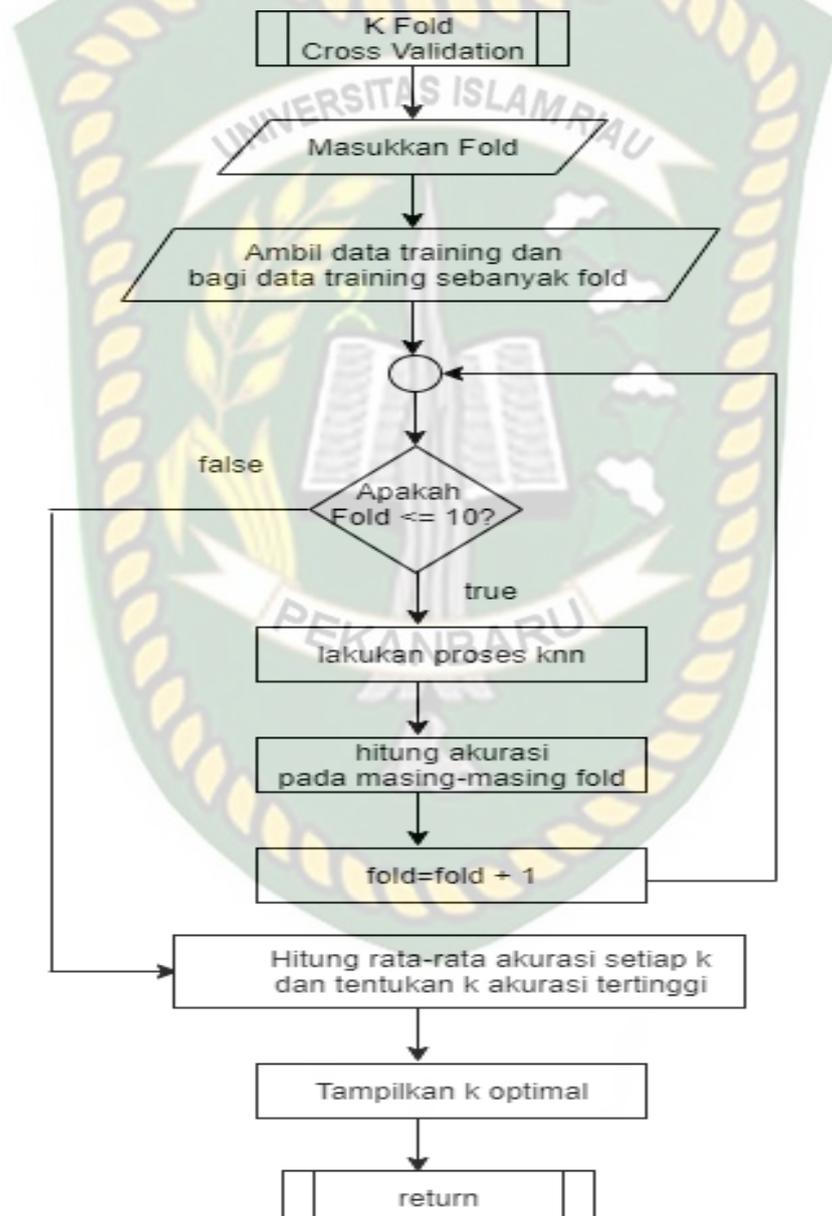
Flowchart perhitungan data testing merupakan alur data program yang merancang input data testing yang akan di lakukan proses perhitungan k-NN dan menghasilkan output berupa hasil klasifikasi. Atribut yang dibutuhkan dalam proses ini adalah *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant* dan nilai k optimal. Kemudian atribut tersebut diolah menggunakan metode k-NN. Adapun desain program *Flowchart* perhitungan data testing dapat dilihat pada gambar 3.18.



Gambar 3.18 Program *Flowchart* Perhitungan Data Testing

5. Program *Flowchart K Fold Cross Validation*

Program *Flowchart* dibawah menjelaskan alur perhitungan dalam mencari nilai k optimal. Adapun desain program *Flowchart K Fold Cross Validation* dapat dilihat pada gambar 3.19.



Gambar 3.19 Program *Flowchart K Fold Cross Validation*

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengujian Hasil

Pengujian pada sistem klasifikasi kualitas retensi surfaktan pada batuan *sandstone* dilakukan terhadap proses data training untuk mendapatkan nilai k-optimal dan proses data testing dengan menggunakan rumus *Euclidean Distance*. Dalam menentukan kriteria untuk kualitas retensi surfaktan dilakukan wawancara dengan instansi yang bersangkutan.

Dalam proses pengujian sistem melewati beberapa tahapan diantaranya yaitu halaman utama sistem, di halaman utama sistem akan menampilkan beberapa menu seperti data *training*, data *testing*, *k-fold cross validation* dan user. Sebelum sistem yang dibangun dipublikasikan, ada beberapa tahapan yang harus dilakukan, hal ini dimaksudkan agar ketika sistem benar-benar sudah dipublikasikan tidak terjadi lagi kesalahan. Dalam pengujian sistem ini dilakukan menggunakan metode *black box* dan pengujian data uji.

4.1.1 Pengujian Akurasi Sistem

Pengujian dibawah ini dilakukan untuk mengukur kinerja klasifikasi kualitas retensi surfaktan pada sistem. Pengukuran ini untuk mengetahui apakah data yang diklasifikasikan ke dalam kelas tertentu secara akurat atau tidak dengan menggunakan 10 data uji. *Configuration matrix* akan menghasilkan nilai akurasi dan *error rate*. Perhitungan akurasi sistem dapat dilihat pada tabel 4.1.

Tabel 4.1 Pengujian Akurasi Sistem

No	Manual	Sistem	Akurasi Manual dan Sistem
1	Tidak Baik	Tidak Baik	Valid
2	Tidak Baik	Tidak Baik	Valid
3	Baik	Tidak Baik	Non Valid
4	Baik	Baik	Valid
5	Baik	Baik	Valid
6	Tidak Baik	Baik	Non Valid
7	Baik	Baik	Valid
8	Baik	Baik	Valid
9	Baik	Baik	Valid
10	Baik	Baik	Valid

Keterangan :

1. Data testing = 10
2. Valid = 9
3. Non Valid = 1

Tabel 4.2 Tabel *Confusion Matrix*

<i>Confusion Matrix</i>	<i>Predicted Class</i>	
	Baik	Tidak Baik
Baik	6	1
Tidak Baik	1	2

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{6+2}{6+2+1+1} \times 100\% \\
 &= \frac{8}{10} \times 100\% \\
 &= 80\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Error Rate} &= \frac{FP+FN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{1+1}{6+2+1+1} \times 100\% \\
 &= \frac{2}{10} \times 100\% \\
 &= 20\%
 \end{aligned}$$

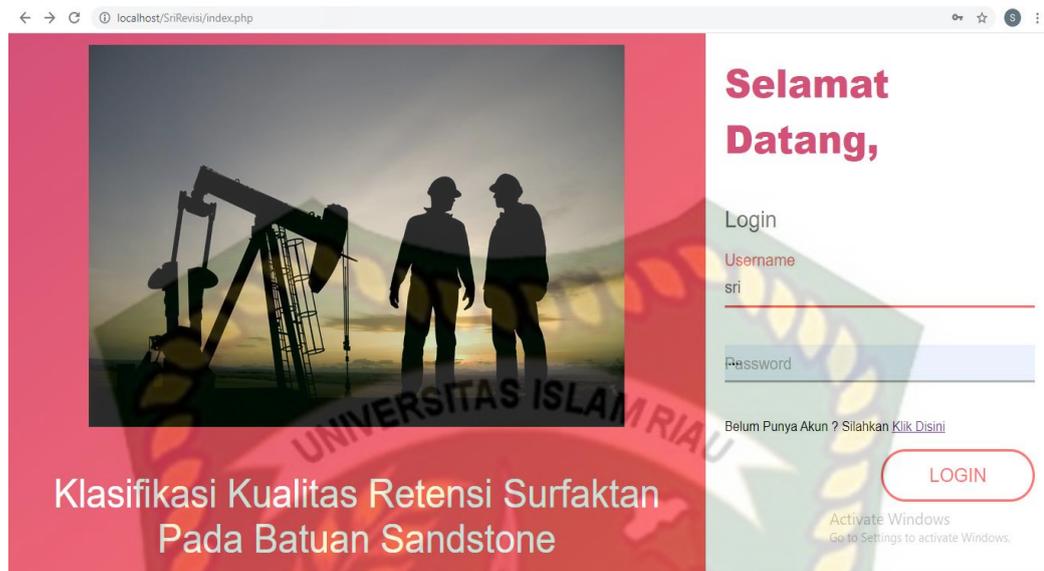
Hasil perhitungan *Confusion Matrix* : untuk akurasi kualitas retensi surfaktan pada batuan *Sandstone* dengan ketetapan akurasi 80 % Valid .

4.2 Pengujian *BlackBox*

Sistem klasifikasi kualitas retensi surfaktan pada batuan *sandstone* ini merupakan hasil dari rancangan *input* atau *output* pada rancangan sebelumnya. Adapun menu-menu yang disediakan dan hasil pengujian *blackbox* ini adalah sebagai berikut :

4.2.1 Pengujian Login

Pada halaman pengujian login ini pengguna akan menginputkan *username* dan *password* untuk masuk kedalam sistem. Untuk melakukan pengujian sistem form login dilihat pada gambar 4.1.



Gambar 4.1 Tampilan *Login*

Apabila data *login* tidak sesuai dengan data pengguna sistem, maka sistem informasi akan mengeluarkan pemberitahuan gagal atas *login* yang dilakukan, kemudian *user* akan diarahkan kembali ke form *login*.

Tabel 4.3 Pengujian *Form Login*

No	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Form <i>login</i>	<i>Username</i> : - <i>Password</i> : - Kemudian klik <i>login</i>	Sistem menolak dengan menampilkan pesan : “ <i>Please Fill Out This Field!</i> ”	[✓] Sesuai [] Tidak Sesuai
		<i>Username</i> : coba (salah) <i>Password</i> : 123 (salah) Kemudian klik <i>login</i>	Sistem menolak dengan menampilkan pesan : “ <i>Username atau Password Salah!</i> ”	[✓] Sesuai [] Tidak Sesuai
		<i>Username</i> : sri <i>Password</i> : 12345 Kemudian klik <i>login</i>	Sistem menerima dan menampilkan halaman menu utama	[✓] Sesuai [] Tidak Sesuai

4.2.2 Pengujian Data User

Pada pengolahan data user ini ada 2 (dua) pengujian yang dilakukan yaitu, penambahan dan edit data pengguna (user). Untuk penambahan data user dapat dilihat pada gambar 4.2.



Daftar Akun

Silakan Di isi

Nama Lengkap
sri rezeki

Username
srirezeki344@gmail.com

Password

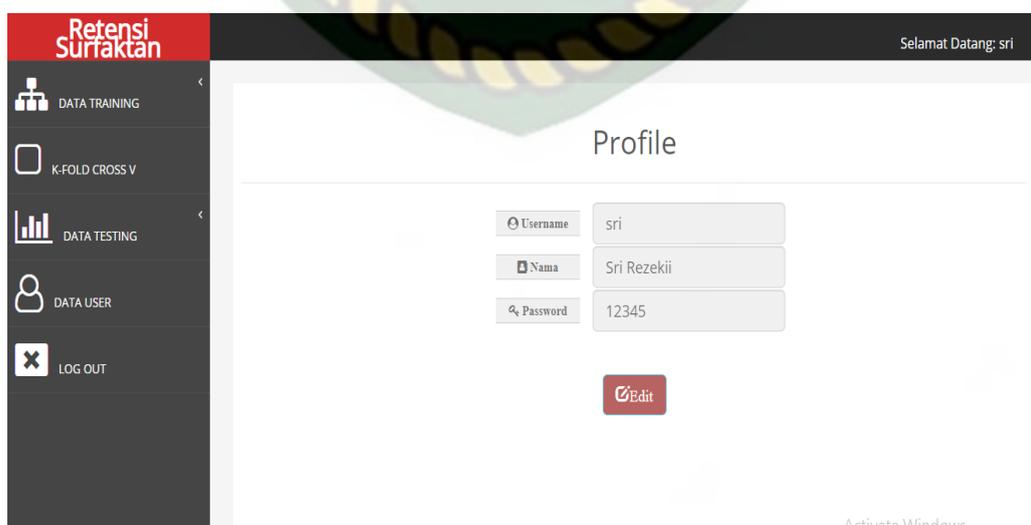
[TAMBAH](#) [KEMBALI](#)

Activate Windows
Go to Settings to activate Windows.

Klasifikasi Kualitas Retensi Surfaktan Pada Batuan Sandstone

Gambar 4.2 Tampilan Tambah Data User

Apabila ada data user yang ingin diubah, user dapat melakukannya pada menu edit user seperti gambar 4.3.



Retensi Surfaktan

Selamat Datang: sri

Profile

Username: sri

Nama: Sri Rezekii

Password: 12345

[Edit](#)

Activate Windows

Gambar 4.3 Tampilan Edit Data User

Adapun penjelasan lengkap pengujian data user dapat dilihat pada tabel 4.4 .

Tabel 4.4 Pengujian Data User

No	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Data user	Memasukkan seluruh data dengan benar, lalu mengklik tombol “Daftar”.	Sistem akan langsung menyimpan data baru dan memasukkannya ke dalam <i>database</i> .	[✓] Sesuai [] Tidak Sesuai
		Memasukkan data dengan salah satu dikosongkan, lalu mengklik tombol “Daftar”.	Sistem menolak dengan menampilkan “ <i>Please fill out this field</i> ” pada kolom yang kosong.	[✓] Sesuai [] Tidak Sesuai
		Mengklik tombol “Edit” pada menu	Akan diarahkan untuk mengedit data, setelah diinputkan maka akan tersimpan di <i>database</i> .	[✓] Sesuai [] Tidak Sesuai

4.2.3 Pengujian Data *Training*

Pada menu data *training* ada 3 (tiga) pengujian yang dilakukan, diantaranya adalah penambahan, edit dan hapus data *training*. Tampilan menu data *training* dapat dilihat pada gambar 4.2 .

No Data	Kabs (mD)	TAN (mg KOH/g)	Temperature (°C)	Co Solvent (wt%)	Polymer Salinity (ppm)	Max Effluent (pH)	Mobility Ratio (dPw/dPpd)	MW Surfaktan (mol)	Retensi Surfaktan	Aksi
1	207	0	25	0.01	11000	7	0.62	559.51	Tidak Baik	Edit Hapus
2	195	0	25	0.01	11000	7.5	0.95	559.51	Tidak Baik	Edit Hapus
3	324	0	25	0.01	15250	6	0.84	559.51	Tidak Baik	Edit Hapus

Gambar 4.4 Tampilan Menu *Data Training*.

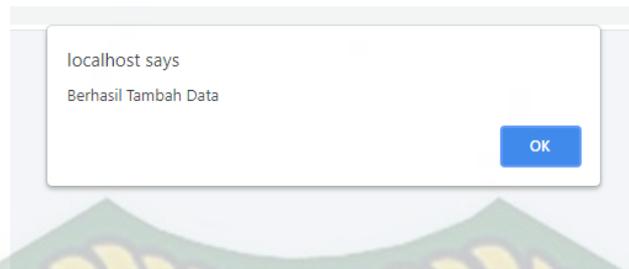
Apabila pengguna ingin menambahkan data training yang baru maka tampilan untuk penambahan data training dapat dilihat pada gambar 4.5

No	Field Name	Value	Unit
1.	No Data	91	
2.	Kabs	323	Md
3.	Tan	3.2	mg KOH/g
4.	Temperature	93	C
5.	Co Solvent	0.002	wt%
6.	Polimer Salinity	4343	ppm
7.	Max Effluent	0.7	pH
8.	Mobility Ratio	23	dPw/dPpd
9.	MW Surfaktan	2324	mol
10.	Retensi Surfaktan	Baik	

Buttons: [Reset](#) [Simpan](#)

Gambar 4.5 Tampilan Tambah *Data Training*

Setelah menginputkan nilai adalah *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant* dan *retensi surfaktan* maka klik tombol simpan maka akan tampil warning pada gambar 4.6



Gambar 4.6 Tampilan Berhasil Tambah Data *Training*

Apabila pengguna melakukan kesalahan dalam menginput data dan ingin mengeditnya, maka tampilan untuk mengedit data training dapat dilihat pada gambar 4.7.

Edit Data Training

1. No Data	1
2. Kabs	207
3. Tan	0
4. Temperature	25
5. Co Solvent	0.01
6. Polimer Salinaty	11000
7. Max Effluent	7
8. Mobility Ratio	0.62
9. MW Surfaktan	559.51
10. Retensi Surfaktan	Tidak Baik

Reset Simpan

Gambar 4.7 Tampilan Edit Data *Training*

Pengujian sistem gambar 4.3 dan gambar 4.4. dapat dijelaskan pada tabel 4.5.

Tabel 4.5 Pengujian Form Data Training

No	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Form Data Training	Mengosongkan inputan data	Sistem menolak dengan menampilkan pesan : <i>“Please Fill Out This Field!”</i>	[✓] Sesuai [] Tidak Sesuai
		Mengisi semua field, Kemudian klik simpan	Data training baru dapat ditambahkan	[✓] Sesuai [] Tidak Sesuai
		Mengedit data yang ada pada data training	Data training dapat di ubah	[✓] Sesuai [] Tidak Sesuai
		Menghapus salah satu data yang ada pada daftar data training	Data training dapat di hapus	[✓] Sesuai [] Tidak Sesuai

4.2.4 Pengujian Data Testing

Pada menu data *testing* ini ada 4 (empat) pengujian yang dilakukan, yaitu penambahan, edit, hapus dan proses data *testing*. Tampilan menu data *testing* dapat dilihat pada gambar 4.8.

The screenshot shows the 'Retensi Surfaktan' web application. The left sidebar contains navigation options: DATA TRAINING, K-FOLD CROSS V, DATA TESTING (highlighted), DATA USER, and LOG OUT. The main content area displays the 'Tabel Data Testing' with a search bar and a table of data. The table has columns for No Data, Kabs (mD), TAN (mg KOH/g), Temperature (°C), Co Solvent (wt%), Polymer Salinity (ppm), Max Effluent (pH), Mobility Ratio (dPw/dPpd), MW Surfaktan (mol), and Aksi. Two records are visible in the table.

No Data	Kabs (mD)	TAN (mg KOH/g)	Temperature (°C)	Co Solvent (wt%)	Polymer Salinity (ppm)	Max Effluent (pH)	Mobility Ratio (dPw/dPpd)	MW Surfaktan (mol)	Aksi
1	478	0.1	100	1	25000	7	0.93	679.91	[Edit] [Hapus] [Tambah]
2	187	0.1	100	1	20000	7	0.2	601.6	[Edit] [Hapus] [Tambah]

Gambar 4.8 Tampilan Menu Data Testing

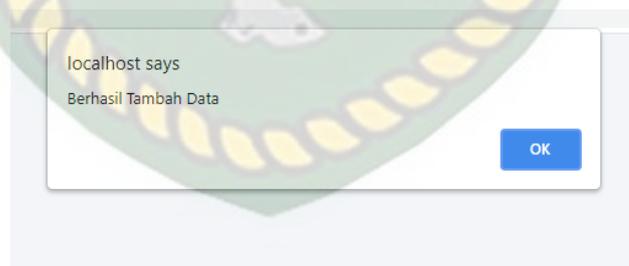
Apabila pengguna ingin menambahkan data testing yang baru maka tampilan untuk penambahan data testing dapat dilihat pada gambar 4.9.

The screenshot shows a web form titled 'Tambah Data' with a 'Data Testing' section. It contains nine input fields with corresponding values and units. At the bottom, there are 'Reset' and 'Simpan' buttons.

No	Parameter	Value	Unit
1.	No Data	11	
2.	Kabs	343	Md
3.	Tan	0.004	mg KOH/g
4.	Temperature	23	C
5.	Co Solvent	0.01	wt%
6.	Polimer Salinity	7	ppm
7.	Max Effluent	7.7	pH
8.	Mobility Ratio	78	dPw/dPpd
9.	MW Surfaktan	5453	mol

Gambar 4.9 Pengujian Tambah Data *Testing*

Setelah menginputkan nilai *Absolute Permeability*, *Total Acid Number*, *Temperature*, *Co-solvent*, *Polimer Salinity*, *Max Effluent pH*, *Mobility Ratio* dan *Molecular weight (MW) Surfactant*. Maka klik tombol simpan maka akan tampil warning pada gambar 4.10 bahwasanya data berhasil disimpan.



Gambar 4.10 Berhasil Tambah Data *Testing*

Apabila pengguna melakukan kesalahan dalam menginput data dan ingin mengeditnya, maka tampilan untuk mengedit data testing dapat dilihat pada gambar 4.11.

Edit Data Testing

1. No Data	1
2. Kabs	478
3. Tan	0.1
4. Temperature	100
5. Co Solvent	1
6. Polimer Salinity	25000
7. Max Effluent	7
8. Mobility Ratio	0.93
9. MW Surfaktan	679.91

Gambar 4.11 Pengujian Edit Data Testing

Selanjutnya pengujian data testing dapat dilakukan jika mengklik tombol “Proses Testing”. Adapun hasil dari proses data testing tersebut dapat dilihat pada gambar 4.12.

Retensi Surfaktan

Selamat Datang: sri

- DATA TRAINING
- K-FOLD CROSS V
- DATA TESTING
- DATA USER
- LOG OUT

Hasil Testing

No	No Data	Kabs	Tan	Temperatur	Co Solvent	Polimer Salinity	Max Effluent	Mobility Ratio	MW Surfaktan	Jarak	Termasuk Retensi?
1	8	478	0.1	100	0	25000	7	0.93	679.91	1	Tidak Baik
2	63	478	0.1	100	0	25000	7	0.93	679.91	1	Tidak Baik
3	42	115	0.5	83	0,001	26000	10	0,07	849.76	1077.46	Baik

RETENSI	
Baik	Tidak Baik
0	3

HASIL NILAI TERTINGGI

Tidak Baik

KESIMPULAN :

Kabs = 478, Tan = 0.1, Temperature = 100, Co Solvent = 1, Polimer Salinity = 25000, Max Effluent = 7, Mobility Ratio = 0.93, MW Surfaktan = 679.91

Masuk Dalam Klasifikasi Retensi= Tidak Baik

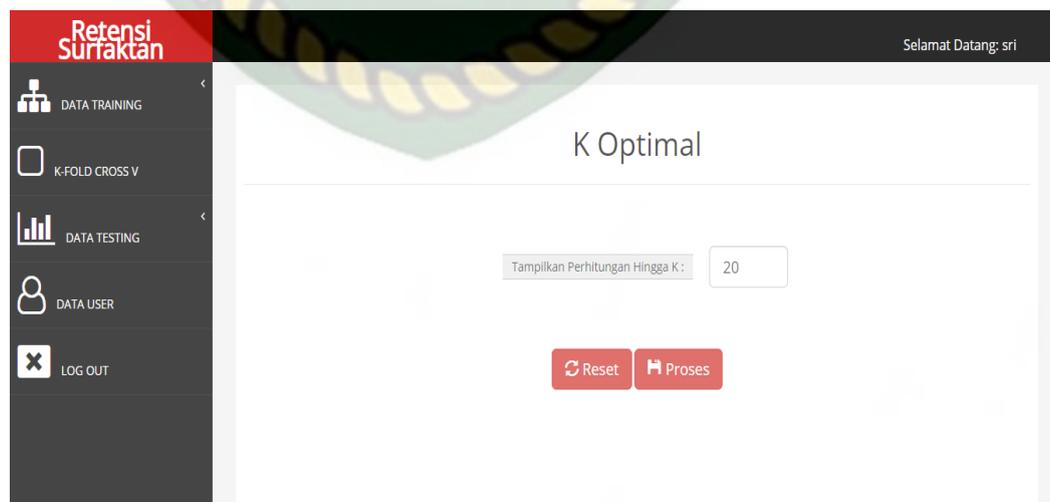
Gambar 4.12 Pengujian data testing

Tabel 4.6 Pengujian Form Data Testing

No	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Form Data Testing	Mengosongkan inputan data	Sistem menolak dengan menampilkan pesan : <i>“Please Fill Out This Field!”</i>	[✓] Sesuai [] Tidak Sesuai
		Mengisi semua field, Kemudian klik simpan	Data testing baru dapat ditambahkan	[✓] Sesuai [] Tidak Sesuai
		Mengedit data yang ada pada data testing	Data testing dapat di ubah	[✓] Sesuai [] Tidak Sesuai
		Menghapus salah satu data yang ada pada daftar data testing	Data testing dapat di hapus	[✓] Sesuai [] Tidak Sesuai
		Mengklik tombol “Proses Testing”	Sistem menampilkan nilai klasifikasi data testing	[✓] Sesuai [] Tidak Sesuai

4.2.5 Pengujian *K Fold Cross Validation*

Pada *k Fold Cross Validation* ini, dilakukan pengujian berupa menginputkan sebanyak *k* yang dibutuhkan yang dapat dilihat pada gambar 4.13.

Gambar 4.13 Pengujian *input* nilai *k*

Dan setelah melakukan penginputan k maka akan tampilkan nilai akurasi k Fold Cross Validation hingga k=15 seperti inputat gambar 4.14.

K FOLD CROSS VALIDATION

Nilai k = 3	
Percobaan	Retensi Surfaktan (%)
Percobaan-1	55.56
Percobaan-2	66.67
Percobaan-3	66.67
Percobaan-4	100
Percobaan-5	77.78
Percobaan-6	100
Percobaan-7	77.78
Percobaan-8	100
Percobaan-9	100
Percobaan-10	77.78
Rata - Rata	82.22
82.22	
Nilai k = 5	
Percobaan	Retensi Surfaktan (%)
Percobaan-1	33.33
Percobaan-2	66.67
Percobaan-3	66.67
Percobaan-4	100
Percobaan-5	88.89
Percobaan-6	100
Percobaan-7	77.78
Percobaan-8	100
Percobaan-9	100
Percobaan-10	77.78

Activa
Go to Se

Gambar 4.14 Pengujian *k Fold Cross Validation*

Pengujian sistem gambar 4.14 dapat dijelaskan pada tabel 4.7.

Tabel 4.7 Pengujian *k Fold Cross Validation*

No	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Form <i>k Fold Cross Validation</i>	Memasukkan nilai k, lalu mengklik tombol "Proses"	Sistem akan langsung melakukan proses perhitungan k optimal	[✓] Sesuai [] Tidak Sesuai
		Mengosongkan data lalu mengklik tombol "Simpan".	Sistem menolak dengan menampilkan pesan : "Please Fill Out This Field!"	[✓] Sesuai [] Tidak Sesuai

4.3 Pengujian Sistem Terhadap Pengguna

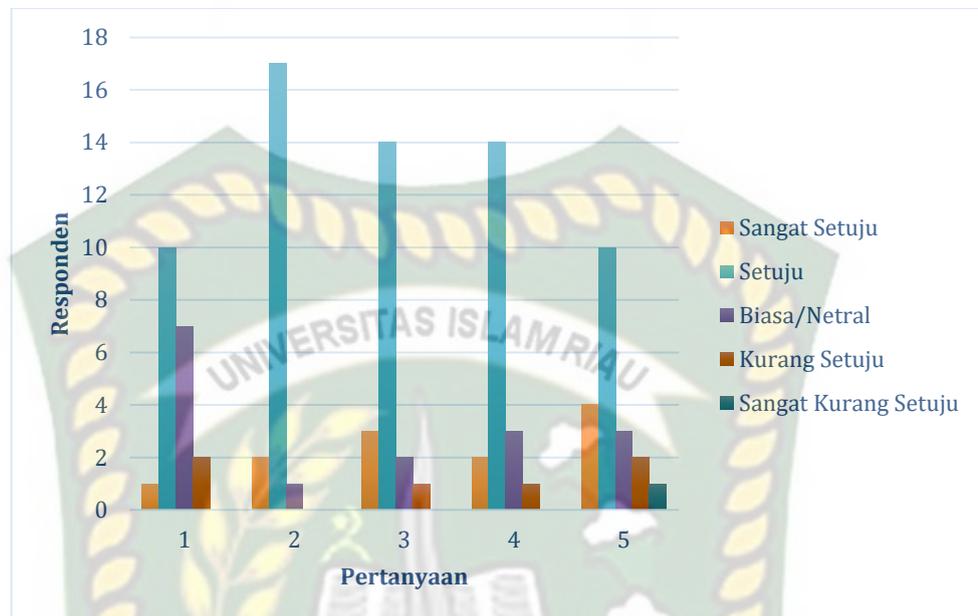
Salah satu pengujian sistem terhadap pengguna yang dilakukan yaitu dengan membagikan kuisioner kepada mahasiswa/i Teknik Perminyakan dan Teknik Informatika sebanyak 20 lembar kuisioner yang berisikan masing-masing 5 pertanyaan. Pengukuran terhadap implementasi sistem ini dilakukan menggunakan Skala Likert dengan penilaian skor 5=Sangat Setuju, 4=Setuju, 3=Biasa, 2=Tidak Setuju dan 1=Sangat Tidak Setuju.

Adapun kelima pertanyaan yang dimaksud adalah sebagai berikut :

1. Apakah klasifikasi kualitas retensi surfaktan pada batuan sandstone ini sesuai dengan kondisi realnya?
2. Apakah sistem ini mudah dimengerti untuk user?
3. Apakah tata letak tampilan atau interface mudah dikenali?
4. Apakah bahasa yang digunakan dalam sistem ini mudah dipahami?
5. Apakah sistem ini dapat memberikan manfaat bagi pengguna?

Dari pertanyaan-pertanyaan tersebut, maka hasil jawaban atau tanggapan dari koresponden terhadap kinerja dari sistem berdasarkan pertanyaan yang diajukan adalah sebagai berikut :

GRAFIK HASIL KUISONER



Gambar 4.14 Grafik Hasil Kuisioner Pengujian Sistem Terhadap Pengguna

Berdasarkan hasil kuisioner tersebut maka dapat disimpulkan bahwa sistem klasifikasi kualitas retensi surfaktan pada batuan *sandstone* dengan menggunakan metode *k-nearest neighbor* ini memiliki persentase sebagai berikut:

Tabel 4.7 Hasil Nilai Persentase Tiap Pertanyaan Kuisisioner

No	Pertanyaan	Jumlah Persentase Koresponden				
		Sangat Setuju	Setuju	Biasa	Tidak Setuju	Sangat Tidak Setuju
1	Apakah klasifikasi kualitas retensi surfaktan pada batuan sandstone ini sesuai dengan kondisi realnya?	1	10	7	2	0
2	Apakah sistem ini mudah dimengerti untuk user?	2	17	1	0	0
3	Apakah tata letak tampilan atau interface mudah dikenali?	3	14	2	1	0
4	Apakah bahasa yang digunakan dalam sistem ini mudah dipahami?	2	14	3	1	0
5	Apakah sistem ini dapat memberikan manfaat bagi pengguna?	4	10	3	2	1
	Jumlah	12	65	16	6	1
	Jumlah Skor	60	260	48	12	1
	Σ Skor Observasi			381		
	Persentase			76,2 %		

Maka secara keseluruhan penilaian kualitas aspek sistem ini dapat dihitung dengan menggunakan perhitungan Skala Likert dengan diperoleh presentase sebagai berikut :

1. Skor maksimal

Skor maksimal adalah skor terbesar pada Skala Likert yang dikalikan dengan jumlah soal, sehingga : $5 \times 5 = 25$

2. Skor yang diharapkan

Skor yang diharapkan adalah skor maksimal yang dikalikan dengan jumlah koresponden, sehingga : $25 \times 20 = 500$

3. Skor observasi

Skor observasi adalah jumlah dari skor masing-masing butir pertanyaan yang dikalikan dengan bobot skor. Adapun skor observasi yaitu sebagai berikut :

$$\begin{aligned} \sum \text{Skor Observasi} = & ((\text{Jumlah jawaban "Sangat Setuju"} * \text{Skor Likert} \\ & \text{"Sangat Setuju"}) + (\text{Jumlah jawaban "Setuju"} * \text{Skor} \\ & \text{Likert "Setuju"}) + (\text{Jumlah jawaban "Biasa"} * \text{Skor} \\ & \text{Likert "Biasa"}) + (\text{Jumlah jawaban "Tidak Setuju"} * \\ & \text{Skor Likert "Tidak Setuju"}) + (\text{Jumlah jawaban} \\ & \text{"Sangat Tidak Setuju"} * \text{Skor Likert "Sangat Tidak} \\ & \text{Setuju"})) \end{aligned}$$

$$\begin{aligned} \sum \text{Skor Observasi} &= ((12 \times 5) + (65 \times 4) + (16 \times 3) + (6 \times 2) + (1 \times 1)) \\ &= ((60) + (260) + (48) + (12) + (1)) \\ &= 381 \end{aligned}$$

4. Perhitungan Persentase

$$\begin{aligned} \text{Persentase} &= \sum \text{Skor observasi} / \text{Skor diharapkan} * 100\% \\ &= 381/500 * 100\% \end{aligned}$$

= 76,2 %

Dari hasil persentase tabel 4.6 di atas dapat disimpulkan hasil pengujian sistem klasifikasi kualitas retensi surfaktan pada batuan *sandstone* dengan menggunakan metode *k-nearest neighbor* yang telah dilakukan kepada 20 koresponden adalah 76,2 % sehingga sistem ini dapat diimplementasikan.



BAB V

KESIMPULAN DAN SARAN

1.1 Kesimpulan

Berdasarkan hasil perancangan sistem klasifikasi metode *K-Nearest Neighbor* dalam menentukan kualitas retensi surfaktan pada batuan *sandstone* serta melakukan serangkaian pengujian pada sistem yang dibangun untuk menguji kemampuan sistem, dapat disimpulkan sebagai berikut :

1. Sistem ini dapat mengklasifikasi kualitas retensi surfaktan pada batuan sandstone yaitu kategori baik dan tidak baik.
2. Tingkat akurasi yang dihasilkan berdasarkan dari hasil pengujian yang dilakukan untuk menentukan kualitas retensi surfaktan pada batuan *sandstone* memiliki tingkat akurasi sebesar 82,22 % dengan nilai $k=3$.
3. Hasil pengujian sistem terhadap pengguna memperoleh persentase kuisisioner tersebut adalah sebesar 76,2 % atau diinterpretasikan sangat baik untuk diimplementasikan.

1.2 Saran

Saran dari penulis untuk sistem klasifikasi kualitas retensi surfaktan pada batuan *sandstone* dengan menggunakan metode *K-Nearest Neighbor* adalah sebagai berikut :

1. Penelitian berikutnya disarankan dapat menggunakan metode lain seperti metode regresi, Support Vector Machine (SVM) dan metode lainnya.
2. Mengembangkan sistem sekarang menjadi sistem aplikasi berbasis *mobile*.

DAFTAR PUSTAKA

- A.S Rosa , dan M.Shalahuddin. 2014. Rekayasa Perangkat Lunak Struktur dan Berorientasi Objek. Bandung : Informatika.
- Abdelgawad, K., & Mahmoud, M., 2014, High-Performance EOR System in Carbonate Reservoirs, *SPE Saudia Arabia Section Technical Symposium and Exhibition*
- Admojoa, Fadhila Tangguh, Ahsanawati.2020. Indonesian Journal of Data and Science. 1(2):34-38
- Ahmadi, M. A. (2014). Developing a Robust Surrogate Model of Chemical Flooding Based on the Artificial Neural Network for Enhanced. Hindawi Journal, 1–9.
- Al-Bahra Bin Ladjamudin. 2006. Analisis dan Desain Sistem Informasi. Graha Ilmu. Yogyakarta
- Alexander F. K. Sibero, 2011, Kitab Suci Web Programming, MediaKom, Yogyakarta
- Aziz, Huzain, Purnawansyah, Farniwati Fattah, Inggrianti Pratiwi Putri.2020. Performa Klasifikasi K-NN Dan *Cross-Validation* Pada Data Pasien Pengidap Penyakit Jantung. ILKOM Jurnal Ilmiah. 12(2): 81-86.
- Banjarsari, Mutiara A., 2015, Pencarian k-Optimal pada Algoritma kNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Berdasarkan IP Sampai Dengan Semester 4, FMIPA Unlam, Banjarbaru.
- Cahyantia, Dewi, lifah Rahmayania, Syafira Ainy Husniara.2020. Indonesian Journal of Data and Science. 1(2): 39-43.
- Chengzhi, B. Y., Bazin, B., Labrid, & Yanli, L. (1988). Reduction of Surfactant Retention with Polyphosphates In Surfactant Flooding Process. Society of Petroleum Engineers Journal.
- Glovesr, C. ., Puerto, M. ., Maerker, J. ., & Sandvik, E. . (1979). Surfactant Phase Behavior and Retention in Porous Media. Society of Petroleum Engineers Journal, 183–193.
- Gorunescu, Florin., 2011, Data Mining: Concepts, Models, and Technique, Verlag, Berlin Heidelberg, Springer.
- Han, J & M Kamber., 2006, Data Mining Concepts and Techniques Second Edition, San Francisco, Morgan Kaufmann.
- Hasanah, Riyan Latifahul, dkk. 2019. Jurnal Techno Nusa Mandiri. Vol.16
- Hermawati Fajar Astuti., 2013, Data Mining, Penerbit Andi, Yogyakarta.
- Iglauer, S., Wu, Y., Shuler, P., Tang, Y., & Iii, W. A. G. (2010). New surfactant classes for enhanced oil recovery and their tertiary oil recovery potential. Journal of Petroleum Science and Engineering, 71, 23–29.
- Kamari, A., Nikookar, M., Sahranavard, L., & Mohammadi, A. H. (2014). Efficient screening of enhanced oil recovery methods and predictive economic analysis. Neural Comput & Applic Journal.
- Kamari, A., Sattari, M., Mohammadi, A. H., & Ramjugernath, D. (2015). Reliable

- method for the determination of surfactant retention in porous media during chemical flooding oil recovery. *FUEL Journal*, 158, 122–128.
- Mandal, A. (2015). Chemical flood enhanced oil recovery. *International Journal of Oil Gas and Coal Technology*, 241–264.
- Mohammed J. Zaki, Wagner Meira JR., “Data Mining and Analysis,” dalam *Data Mining and Analysis Fundamental Concepts and Algorithms*, New York, Cambridge University Press, 2014, p. 1.
- Novosad, J. (1982). Surfactant Retention In Berea Sandstone- Effects of Phase Behavior and Temperature. *Society of Petroleum Engineers Journal*, 962–970.
- Pandie, Emerensye S Y., 2012, Implementasi Algoritma Data Mining K-Nearest Neighbor (KNN) Dalam Pengambilan Keputusan Pengajuan Kredit. Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana, Kupang
- Park, S., Lee, E. S., Rosli, W., & Sulaiman, W. R. W. (2015). Adsorption Behaviors of Surfactants for Chemical Flooding in Enhanced Oil R
- Ridha Muldina Negara, R. T. (2017). Analisis Simulasi Penerapan Algoritma OSPF Menggunakan RouteFlow pada Jaringan Software Defined Network (SDN). *Jurnal Infotel*, 77.
- Rizal, Azwar., 2013, Perbandingan Performa antara Imputasi Metode Konvensional dan Imputasi dengan Algoritma Mutual Nearest Neighbor, Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember. Surabaya.
- Satzinger, Gary., Rosenblatt, Harry J., 2012, *System Analysis and Design in a Changing World*, Boston, MA, Course Technology
- Sheats, W. Brad dan Norman C. Foster, 1997. *Concentrated Products from Methyl Ester Sulfonates*.
- Sheng, J. (2011). *Modern Chemical Enhanced Oil Recovery*. Elsevier Inc. Sheng, J. (2015). Status of surfactant EOR technology. *Journal of Petroleum*, 1, 97–105.
- Thomas, S. (2007). Enhanced Oil Recovery – An Overview. *Oil & Gas Science and Technology Journal*, 63, 9–19.
- Wang, F. H. . (1993). Effects of Reservoir Anaerobic , Reducing Conditions on Surfactant Retention in Chemical Flooding. *Society of Petroleum Engineers Journal*, 108–116.
- Yassin, M. R., Arabloo, M., Shokrollahi, A., & Mohammad, A. H. (2013). Prediction of Surfactant Retention in Porous Media : A Robust Modelling Approach. *Journal of Dispersion Science and Technology*, 1–40.