

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Fokus penelitian ini adalah pembuatan sistem pemantauan dan *visibility* yang mencakup hasil analisis *service impact* serta gambaran untuk setiap *tower* PT. XYZ menggunakan Grafana. Penelitian ini memiliki tujuan agar dapat memberikan analisis *user service impact* berdasarkan metrik KPI serta memberikan visualisasi keseluruhan untuk setiap *tower* PT. XYZ yang tersebar di Indonesia. Sistem ini juga akan memberikan pemantauan *near real – time* untuk setiap *alarm* maupun kendala yang terjadi dalam sebuah *tower*. Analisis *network service impact* akan disesuaikan dengan metode *classification* yang dapat membantu memberikan label untuk setiap masalah yang ditemukan berdasarkan metrik KPI yang digunakan. Sistem yang akan dibuat akan mengimplementasikan konsep sistem berbasis peta sebagai *output* atau bentuk akhir dari sistem ini. Data yang akan digunakan akan disiapkan menggunakan Pentaho Data Integration (PDI) serta analisis *classification* akan digunakan menggunakan Python. Bentuk data yang disediakan akan bervariasi disesuaikan dengan data yang didapatkan dari PT. XYZ. Bentuk data yang akan digunakan dapat dilihat pada gambar 3.1 yang menunjukkan bentuk data akhir yang digunakan dalam pembuatan sistem yang akan dirancang.

Date	Network	T	mcc	mnc	Operator	activeTest	avgDRate	avgLatency	avgRsp	avgRssi	avgSnr	avgLRate	cellid	deviceId	eNodeB	gsmStatus	device_latitude	device_longitude	site_id	cell_id	latnum	site_longitude	site_latitude	location	antenna_height	Impact
12/31/2023 0:00	LTE	510-35	XYZ	Quick Test	37.15715	178	-38	-12	-14	11.13085	11.33411	82	DC0000000	970540	On	-6.51119	106.90206	G_BGR252	82	150	106.9018	-6.51029	OUTDOOR	30	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Quick Test	97.17993	105	-63	-4	-47	13.11801	18.08105	83	DC0000000	300543	On	0.881401	102.1362	F_PKU054	83	150	102.134	0.881565	OUTDOOR	49	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Quick Test	97.17993	105	-63	-4	-47	13.11801	18.08105	83	DC0000000	300543	On	0.881401	102.1362	F_PKU054	83	340	102.0424	0.738997	OUTDOOR	67	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	55.74738	191	-133	-2	-83	-8.84955	37.79693	82	DC0000000	338457	On	-6.80791	107.0621	G_BOG525	82	160	107.0598	-6.79149	OUTDOOR	47	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	36.99623	905	-54	-4	-97	-4.3584	90.98419	88	DC0000000	890450	On	-8.18222	100.0727	F_P00020	88	15	100.0737	-8.18522	OUTDOOR	49	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	79.05883	175	-121	-19	-73	-9.4367	17.78268	88	DC0000000	830090	On	-6.50072	106.8827	F_BGR007	88	50	106.884	-6.50439	OUTDOOR	29	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	98.25492	182	-58	-8	-27	11.06845	59.53013	82	DC0000000	860744	On	-4.97418	105.4813	F_LMP072	82	150	105.4807	-4.9736	OUTDOOR	51	Good cover	
12/31/2023 0:00	LTE	510-10	XYZ	Quick Test	7.561614	3120	-64	-15	-12	-6.49934	19.46683	83	DC0000000	330887	On	-6.70481	107.2726	G_BDG488	83	320	107.2749	-6.22179	OUTDOOR	50	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	45.00329	84	-34	-18	-26	17.34168	54.45899	82	DC0000000	890876	On	-2.80435	103.9354	F_PLB0876	82	195	103.9357	-2.80222	OUTDOOR	40	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	96.90878	34	-120	-7	-51	7.81816	76.93248	88	DC0000000	990008	On	-6.08353	106.7934	G_JKT3400	88	355	106.7931	-6.08359	OUTDOOR	20	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	7.80975	122	-54	-10	-102	15.21719	48.81877	88	DC0000000	290269	On	-3.31095	114.8521	G_BM0106	88	140	114.852	-3.30968	OUTDOOR	27	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Quick Test	55.34927	2992	-72	-21	-26	-2.74845	0.51055	83	DC0000000	300047	On	0.625062	101.5714	F_PKU002	83	235	101.5715	0.625109	OUTDOOR	37	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	106.4306	146	-113	-21	-39	4.228457	66.37538	88	DC0000000	370373	On	-6.36875	107.116	G_BGR437	88	40	107.0599	-6.22117	OUTDOOR	40	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	106.4306	146	-113	-21	-39	4.228457	66.37538	88	DC0000000	370373	On	-6.36875	107.116	G_BGR437	88	40	106.815	-6.48192	OUTDOOR	40	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	106.4306	146	-113	-21	-39	4.228457	66.37538	88	DC0000000	370373	On	-6.36875	107.116	G_BGR437	88	40	107.1179	-6.36892	OUTDOOR	40	Bad cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	111.9977	151	-85	-7	-32	-9.51195	61.3423	83	DC0000000	370373	On	-6.35783	107.1215	G_BGR437	83	270	107.0599	-6.22117	OUTDOOR	35	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	111.9977	151	-85	-7	-32	-9.51195	61.3423	83	DC0000000	370373	On	-6.35783	107.1215	G_BGR437	83	270	106.815	-6.48192	OUTDOOR	35	Good cover	
12/31/2023 0:00	LTE	510-35	XYZ	Full Test	111.9977	151	-85	-7	-32	-9.51195	61.3423	83	DC0000000	370373	On	-6.35783	107.1215	G_BGR437	83	270	107.1179	-6.36892	OUTDOOR	35	Good cover	

Gambar 3. 1 Tampilan Dataset Yang akan digunakan

3.2 Metode Penelitian

Penelitian ini bersifat kuantitatif dikarenakan melakukan pembuatan sistem pemantauan yang didasari oleh data yang diperoleh dari PT. XYZ Penelitian ini bersifat kuantitatif dikarenakan data yang akan digunakan bervariasi sehingga perlu memahami seluruh data yang akan digunakan selain itu juga akan melakukan proses

ETL terhadap data yang akan digunakan. Data yang akan digunakan untuk analisis *machine learning* akan bersifat numerik sedangkan data untuk sistem berbasis peta akan memiliki nilai geografis yang akan digunakan dalam menentukan lokasi untuk setiap *tower* dan *user location*.

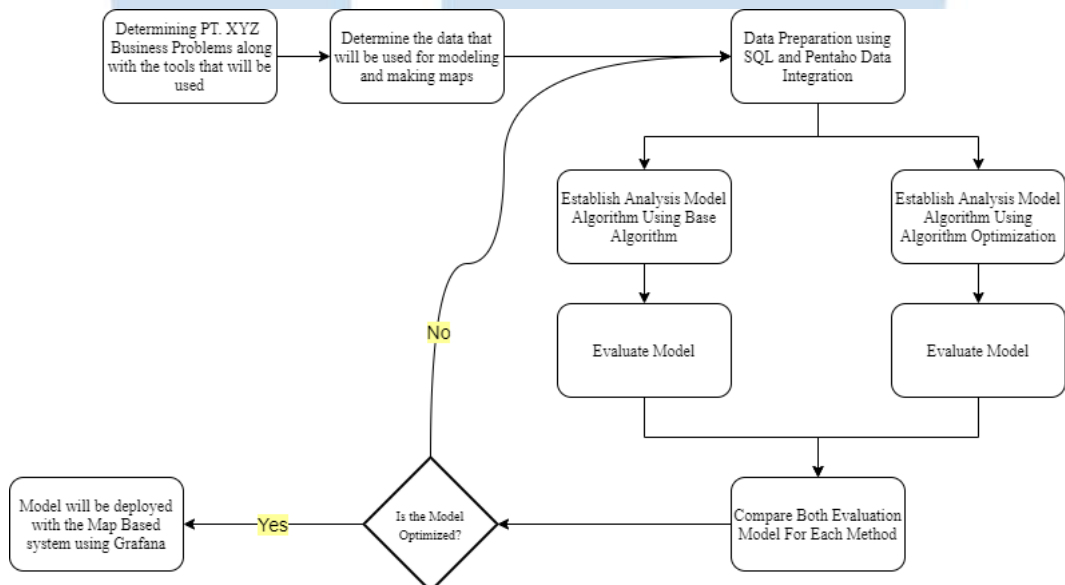
3.2.1 Alur Penelitian

Penelitian ini akan dilakukan dengan metode *cross-industry process for data mining* (CRISP-DM), pada tabel 3.1 menunjukkan perbandingan antara setiap metode *data mining* yang populer. Pemilihan CRISP-DM dikarenakan metode yang fleksibel terhadap berbagai proyek *data mining* selain itu dalam penelitian ini juga memiliki tahap *deployment* atau penerapan model yang akan dilakukan. Seluruh tahapan dari CRISP-DM diterapkan sebagai dasar dari penelitian dikarenakan sudah dapat menjelaskan alur penelitian yang akan dilakukan dalam penelitian ini. Pada gambar 3.2 menunjukkan alur penelitian yang disesuaikan dengan tahapan dari CRISP-DM.

Tabel 3. 1 Perbandingan Framework Data Mining

Framework	Fokus Utama	Kelebihan	Kekurangan
CRISP-DM	Keseluruhan Siklus Proyek	<ul style="list-style-type: none"> • Fleksibel dan cocok untuk berbagai proyek <i>data mining</i> dan <i>machine learning</i>. • Mendukung dokumentasi yang luas • Seluruh fase dapat mengubah masalah dalam proses <i>data mining</i>. 	<ul style="list-style-type: none"> • Membutuhkan waktu dan sumber daya yang signifikan karena akan melibatkan beberapa tahap iterasi. • Memiliki proses yang panjang dan kompleks.
KDD	Penemuan Pengetahuan Data	Memiliki proses yang lebih luas daripada <i>data</i>	Tidak memiliki dukungan sekuat CRISP-DM.

		<i>miningi</i> , mencakup seluruh siklus penemuan pengetahuan dari <i>database</i>	
SEMMA	Pemodelan Data	<i>Framework</i> ini lebih cocok membangun model dalam lingkungan <i>software</i> SAS	Kurang digunakan diluar penggunaan SAS, selain itu tidak sekomprensif CRISP-DM



Gambar 3. 2 Alur Penelitian

3.2.2 Metode Pengembangan Sistem

Berdasarkan alur penelitian, berikut metodologi *framework* CRISP-DM diterapkan pada penelitian ini. Berikut merupakan penjelasan mengenai tahap penelitian menggunakan CRISP-DM sebagai berikut:

1. *Business Understanding* – Dalam tahap ini melakukan identifikasi masalah yang dihadapi oleh PT. XYZ di mana membutuhkan sistem yang dapat memberikan *visibility* untuk *tower* PT. XYZ serta membantu analisis *service impact* dari sisi pengguna untuk layanan jaringan PT.

XYZ. Selain itu tahap ini juga dilakukannya pemahaman mengenai data yang akan digunakan untuk analisis dan pembuatan sistem.

2. *Data Understanding* – Pada tahap ini akan menggunakan 3 *dataset* yakni *nv_active*, *sector_detail*, dan *site_detail*. *Dataset* yang digunakan untuk model akan diambil dari *nv_active*, sedangkan data untuk sistem peta akan menggunakan hasil gabungan tiga *dataset* menjadi *dataset nv_analytics*.
3. *Data Preparation* – Pada tahap ini untuk data *nv_active* akan melakukan proses transformasi data menjadi *dataset* yang akan digunakan untuk *modeling*. Pada gambar 3.3 menunjukkan kolom yang akan digunakan untuk *modeling data* yang terdiri dari 7 kolom *feature* dan 1 kolom *target* “**Impact**” yang sudah ditetapkan oleh pihak perusahaan berdasarkan metrik KPI yang digunakan.

avgDLRate	avgLatency	avgRsrp	avgRsrq	avgRssi	avgSinr	avgULRate	Impact
37.15715	178	-38	-12	-14	11.13085	11.33411	Good coverage and high interference
97.17993	105	-63	-4	-47	13.11801	18.08105	Good coverage and high interference
97.17993	105	-63	-4	-47	13.11801	18.08105	Good coverage and high interference

Gambar 3. 3 Data yang digunakan untuk model machine learning

4. *Modeling* – Dalam tahap ini akan melakukan proses pemodelan data yang sudah difilter. Analisis akan dilakukan dengan metode klasifikasi di mana model *machine learning* akan digunakan untuk membantu prediksi kolom *Impact* yang dapat dilihat pada gambar 3.3. Tahap analisis akan dibantu dengan menggunakan algoritma *Decision Tree* dan *Random Forest*.
5. *Evaluation* – Model dan sistem sudah dibuat dan sudah disesuaikan dengan keinginan dari perusahaan. Terkait model akan melakukan beberapa evaluasi sebelum di impelentasikan ke dalam sistem untuk menggunakan model yang paling optimal. Evaluasi dapat dilakukan dengan melakukan berbagai uji validasi model seperti *cross validation* dan validasi manual menggunakan data buatan sendiri.
6. *Deployment* – Model yang sudah dibuat akan diterapkan menggunakan *joblib* pada sistem pemantauan dan visibilitas berbasis peta yang sudah dibuat agar dapat memberikan gambaran terkait *network problem* yang

dialami. Model yang akan digunakan dalam sistem ini dipilih berdasarkan uji validasi yang paling tinggi. Sistem berbasis peta ini dibuat menggunakan Grafana untuk menyesuaikan aplikasi dan *software* yang digunakan oleh PT. XYZ.

3.3 Teknik Pengumpulan Data

3.3.1 Sumber Data

Pengumpulan data merupakan tahap penting dalam proses penelitian kuantitatif karena dapat mempengaruhi alur serta hasil penelitian. Teknik pengumpulan data merujuk pada metode yang digunakan untuk mengumpulkan data atau informasi yang diperlukan untuk penelitian [36]. Penelitian ini akan menggunakan data *network* yang terdiri dari metrik KPI untuk pembuatan model dan variabel geografis untuk teknologi sistem berbasis peta. Data *network* berasal dari PT. XYZ dan bersifat konfidensial dikarenakan mencakup beberapa data perusahaan sehingga membutuhkan izin dari pihak perusahaan untuk menggunakan data sebagai objek utama dalam penelitian. Data yang akan ditunjukkan berupa data *dummy* atau tiruan dari data asli untuk menjaga kerahasiaan informasi dari pihak perusahaan.

3.3.2 Populasi dan Sampel

Populasi merupakan hal penting dalam proses penelitian. Populasi objek dan subjek dalam penelitian yang memiliki karakteristik tertentu [37]. Penelitian ini memiliki populasi merupakan data *network* pada tahun 2024 yang memiliki rentang waktu 1 Januari 2024 sampai dengan 9 Februari 2024. Sampel merupakan bagian dari jumlah data yang dimiliki populasi [38]. Penelitian ini akan menggunakan teknik *non-probability sampling* yang berarti pengambilan sampel tidak memberikan kesempatan yang sama untuk setiap unsur dalam populasi. Penelitian ini akan menggunakan teknik *purposive sampling* yang merupakan bagian dari *non-probability sampling* yang berarti sampel akan diambil berdasarkan kriteria dan kebutuhan penelitian dan tidak secara acak [39]. Pemilihan teknik *purposive sampling* untuk mencegah adanya *system crash* yang disebabkan karena pemuatan data yang terlalu banyak di dalam peta interaktif yang akan dibuat. Sampel yang diambil akan disesuaikan

dengan kebutuhan dan kriteria penelitian yang sudah dapat memberikan informasi terkait sistem yang ingin dirancang.

3.4 Variabel Penelitian

Variabel penelitian merupakan segala sesuatu yang akan menjadi objek pengamatan dalam penelitian. Berdasarkan hubungannya variabel penelitian akan terbagi menjadi dua yakni variabel independen dan variabel dependen.

1. Variabel Independen

Variabel independen merupakan tipe variabel yang tidak bergantung dengan variabel lainnya atau variabel ini dapat berdiri sendiri. Variabel independen juga dapat memengaruhi variabel dependen lainnya. Dalam penelitian ini yang termasuk dari variabel independen merupakan variabel geografis seperti *latitude*, dan *longitude*.

2. Variabel Dependen

Variabel dependen merupakan tipe variabel yang sangat bergantung terhadap variabel independen. Dalam penelitian ini terdapat beberapa data yang masuk ke dalam variabel dependen yaitu sektor setiap *site* arahnya akan dipengaruhi oleh variabel geografis.

3.5 Teknik Analisis Data

Teknik analisis data akan menggunakan berbagai aplikasi dan *software* yang akan disesuaikan dengan *tools* yang digunakan oleh PT. XYZ. Dalam tahap *data preparation* akan dibantu dengan menggunakan Pentaho Data Integration (PDI), aplikasi ini akan digunakan untuk membantu melakukan filter data atau penghapusan *null value* atau *noise* yang berlebihan dalam dataset. PDI digunakan dikarenakan merupakan aplikasi utama yang digunakan oleh PT. XYZ dalam melakukan proses *extract, transform, and load*.

Dalam proses *modeling* akan menggunakan Google Colab sebagai *coding environment* untuk melakukan analisis *machine learning*. Pemilihan Google Colab dikarenakan merupakan aplikasi berbasis web selain itu seluruh *libraries* umum yang digunakan dalam analisis *machine learning* sudah dapat langsung digunakan tanpa perlu melakukan instalasi *libraries* terlebih dahulu, selain itu dengan

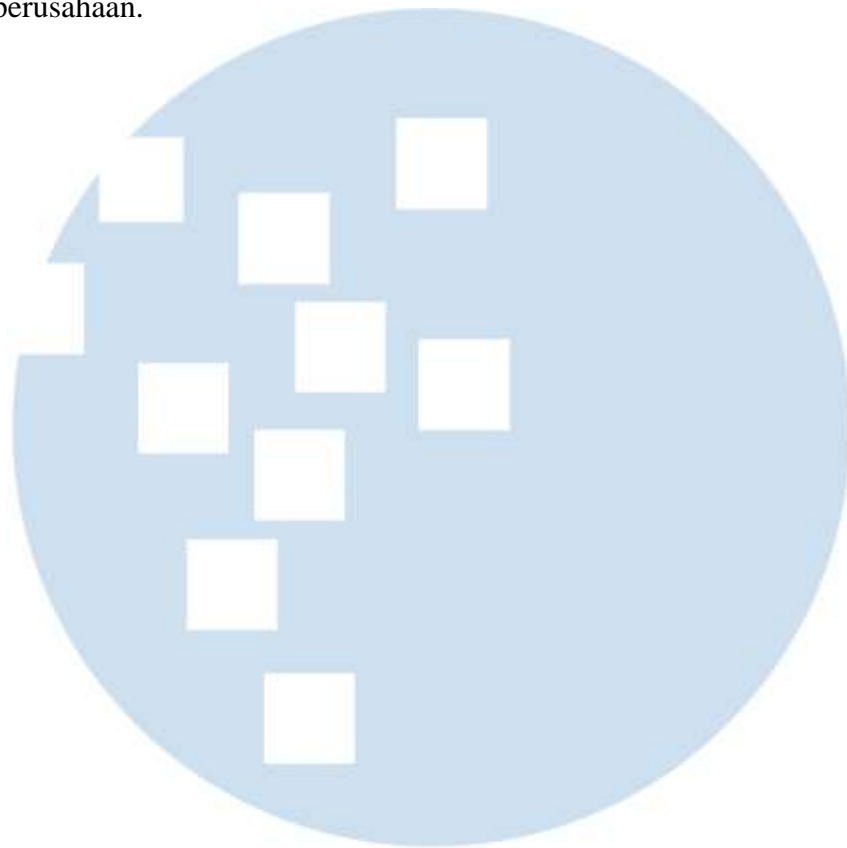
menggunakan Google Colab proses eksekusi kode tidak terlimitasi oleh *hardware* yang digunakan melainkan menggunakan *hardware* yang disediakan oleh pihak Google. Selain Google Colab dalam tahap modeling juga menggunakan bahasa pemrograman Python pada tabel 3.2 menunjukkan tabel perbandingan antara bahasa pemrograman Python dan R yang merupakan bahasa pemrograman statistik populer. Pemilihan Python dikarenakan merupakan bahasa pemrograman statistik yang digunakan oleh pihak PT. XYZ selain itu pemilihan python juga didukung dengan penggunaan model prediksi yang akan digunakan dalam sistem pemantauan peta yang akan dibuat.

Tabel 3. 2 Tabel Perbandingan Python dan R

Kategori	Python	R
Kegunaan	Dapat digunakan untuk <i>data science</i> dan pengembangan <i>software</i>	Digunakan untuk analisis statistik
<i>Library</i>	Terdapat berbagai <i>library</i> yang dapat mendukung analisis <i>data science</i>	<i>Library</i> R selain mendukung analisis <i>data science</i> juga memberikan beberapa <i>library</i> yang mendukung teknik grafis.
Kelebihan	Python memiliki sintaks yang mudah dipahai serta dapat digunakan di luar analisis data seperti pengembangan <i>software</i>	R dirancang untuk analisis statistik dan memiliki IDE sendiri sehingga akan mendukung proses analisis data
Kekurangan	<i>Library</i> Python tidak selengkap R dalam melakukan analisis data serta <i>library</i> visualisasi yang tidak sebaik R	R cenderung lambat dan tidak dapat menganalisa multivariabel dan merupakan bahasa pemrograman yang tergolong sulit dipelajari.

Pada tahap *deployment* model yang sudah dibuat akan diterapkan menggunakan *library* *joblib* agar dapat digunakan untuk melakukan prediksi pada data lain. Proses prediksi secara otomatis ini akan dibantu menggunakan PDI dengan memasukkan data ke dalam database. Setelah data masuk ke dalam *database* model akan di terapkan pada sistem pemantauan berbasis peta menggunakan Grafana sebagai *software* visualisasi utama yang digunakan. Pemilihan Grafana dibandingkan dengan program visualisasi lainnya dikarenakan Grafana merupakan *software open-source* yang memiliki berbagai *plugin* untuk membantu visualisasi data. Grafana juga merupakan *tools* yang digunakan oleh PT. XYZ oleh karena itu sistem

ini dirancang menggunakan Grafana untuk menyesuaikan *tools* yang digunakan pihak perusahaan.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA