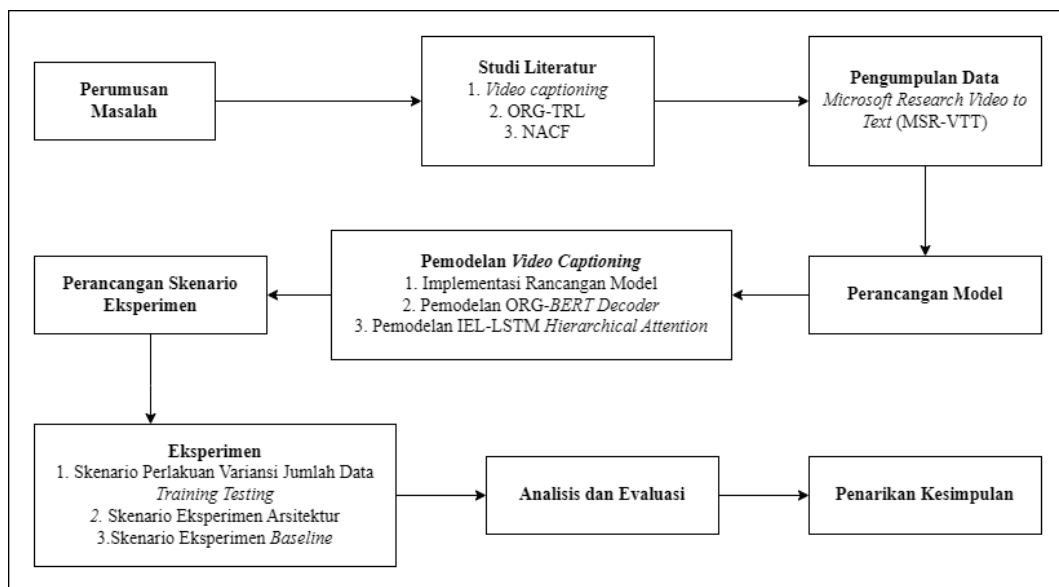


## BAB III METODE PENELITIAN

Bab ini akan menjelaskan tentang metode penelitian, dimulai dari desain penelitian, alat dan bahan penelitian, desain model, diagram alir, dan teknik pengujian sistem.

### 3.1 Desain Penelitian

Desain penelitian adalah kerangka kerja yang digunakan untuk melakukan penelitian. Pada bagian ini penulis akan memaparkan kerangka kerja dimulainya penelitian hingga selesainya penelitian. Gambar 3.1 menunjukkan alur dari penelitian yang akan dilakukan. Secara garis besar desain penelitian meliputi perumusan masalah, studi literatur, pengumpulan data, perancangan model, pemodelan *video captioning*, perancangan skenario eksperimen, eksperimen, analisis dan evaluasi, dan penarikan kesimpulan.



Gambar 3.1 Desain Penelitian yang akan dilakukan

Pada bagian selanjutnya akan dijelaskan secara rinci tentang desain penelitian yang akan dilakukan berdasarkan Gambar 3.1.

#### 3.1.1 Perumusan Masalah

Penelitian dimulai dengan menentukan topik atau perumusan masalah yang akan diselesaikan. Perumusan masalah meliputi upaya memahami latar belakang masalah dan menentukan solusi terhadap masalah tersebut. Kedua hal ini dapat

diperoleh dari literasi penelitian-penelitian terdahulu. Tahap perumusan masalah merupakan tahap yang paling penting karena mendasari hal yang dilakukan selama penelitian.

Perumusan masalah merupakan tahapan penting karena menjadi acuan terhadap hal yang dilakukan selama penelitian. Perumusan masalah memuat identifikasi spesifik dari masalah yang akan diteliti. Hasil identifikasi ini dibuat menjadi poin-poin rumusan masalah dan dijadikan sebagai topik pembahasan selama penelitian dilakukan. Poin-poin tersebut juga akan menjadi dasar penulisan deskripsi tujuan dan manfaat.

Deskripsi tujuan dan manfaat penelitian merupakan hasil yang diperoleh dari tahap perumusan masalah. Tujuan penelitian umumnya mencerminkan hal yang akan diselesaikan dari rumusan masalah. Manfaat penelitian memberikan gambaran kontribusi dari hasil penyelesaian penelitian ini kepada masyarakat.

### 3.1.2 Studi Literatur

Akar dari rumusan masalah dipahami pada tahap studi literatur. Tahap ini dimulai dengan mengumpulkan teori pendukung yang bersumber dari jurnal, buku, *e-book*, dan sumber lainnya. Setelah itu, peneliti mengkaji penelitian-penelitian terdahulu untuk mempelajari bagaimana persoalan yang sedang diteliti diselesaikan. Berdasarkan hasil kajian tersebut, peneliti dapat mengetahui teori atau konsep yang diperlukan untuk melaksanakan penelitian. Lebih jauh lagi, hasil kajian ini membantu peneliti mengetahui wilayah kontribusi peneliti terhadap bidang yang sedang diteliti. Kontribusi ini dapat berupa pengembangan solusi dari penelitian lain atau penemuan solusi baru untuk masalah yang sama.

Literasi yang dikumpulkan dan dipelajari oleh peneliti berkenaan dengan topik *video captioning*, *convolutional neural network*, *Object Relational Graph (ORG)*, *long short-term memory (LSTM)*, *autoregressive decoding*, *non-autoregressive decoding*, dan topik-topik pendukung lainnya. Literasi ini dijadikan sebagai dasar untuk menjawab rumusan masalah dan sebagai referensi untuk memahami metode yang digunakan dalam penelitian.

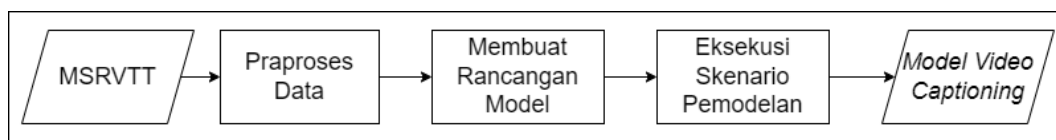
### 3.1.3 Data

Data pada penelitian ini diperoleh melalui pengunduhan data publik. Data publik ini bernama *Microsoft Research Video to Text (MSRVTT)* dan dapat

diunduh pada link <https://www.kaggle.com/datasets/vishnutheepb/msrvtt>. Data ini diperoleh dengan metode *web crawl* pada *search engine* video komersil (Xu dkk., 2016). Secara umum data ini terdiri dari klip video, *caption* terkait video, dan informasi tambahan lainnya. Data ini dibagi menjadi tiga himpunan yang didasarkan dari jumlah klip video yang ada di dataset. Himpunan tersebut adalah *training* dengan jumlah data sebanyak 6.513, *validation* sebanyak 497 data, dan *testing* sebanyak 2.990 data.

### 3.1.4 Perancangan Metode

Perancangan metode merupakan tahap pembuatan rancangan metode yang digunakan untuk menyelesaikan rumusan masalah dari topik penelitian. Rancangan metode ini umumnya diperoleh dari hasil kajian yang telah dilakukan pada tahap studi literatur. Rancangan metode ini dapat berupa pengembangan rancangan metode lain dari penelitian terdahulu atau dapat berupa temuan rancangan metode baru.



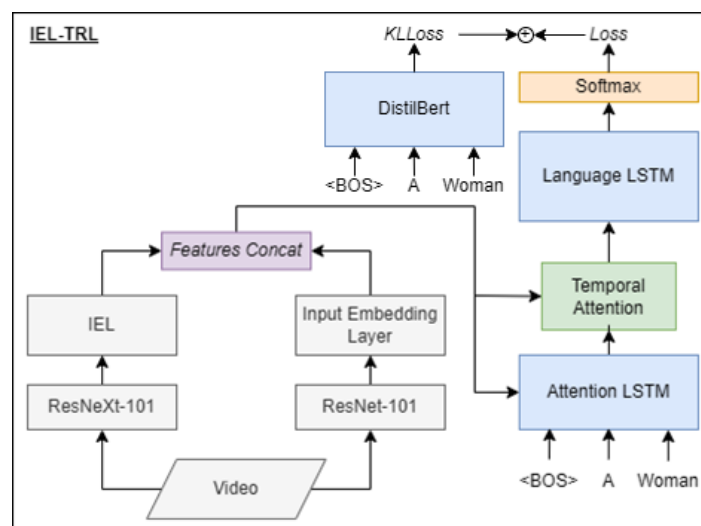
Gambar 3.2 Rancangan metode

Perancangan metode pada penelitian ini terdiri dari tiga tahap utama dan dapat dilihat pada Gambar 3.2. Tahap pertama merupakan praproses data MSRVT. Praproses data dilakukan oleh peneliti terhadap data klip video dan data *caption*. Praproses ini dilakukan agar data MSRVT sesuai dan dapat digunakan untuk *training*. Tahap kedua adalah merancang model yang diajukan. Tahap ketiga adalah melakukan skenario pemodelan merujuk pada rancangan model yang diajukan.

Tahap pertama yaitu praproses data dilakukan terhadap data klip video dan data *caption*. Fitur data klip video dibagi ke dalam tiga fitur di antaranya *image feature*, *motion feature*, dan *object feature*. *Image feature* diperoleh dengan menggunakan model *convolutional neural network* (CNN) ResNet dan InceptionResNetV2 yang sudah dilatih pada dataset ImageNet. *Motion feature* diekstrak menggunakan model 3D CNN seperti ResNeXt yang dilatih pada dataset Kinetics. *Object Feature* diambil menggunakan model *object detection* Faster-

RCNN yang dilatih pada dataset COCO. Semua fitur ini adalah representasi dari klip video dan merupakan masukan untuk model nantinya. Setelah praproses data video dilanjutkan ke praproses data *caption*.

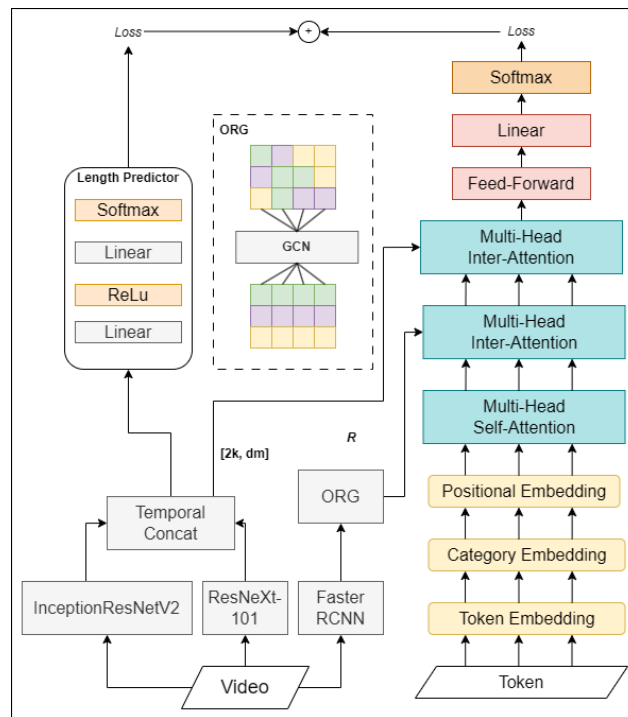
Praproses data *caption* memuat beberapa proses antara lain proses memotong kalimat yang memiliki panjang lebih dari 24 kata, merubah setiap kata ke *lowercase*, dan menghilangkan setiap tanda baca. Kata dimasukkan ke dalam *vocabulary* apabila kata tersebut muncul minimal sebanyak dua kali. Selanjutnya, setiap kata ditokenisasi menggunakan *vocabulary* yang sudah dibuat. Kata yang tidak ada di *vocabulary* akan dirubah menjadi token spesial seperti “[UNK]” jika kata tersebut tidak ada di *vocabulary*. Setiap kalimat diberi token spesial di awal dan akhir kalimat. Token spesial itu adalah “[BOS]” dan “[EOS]”. Token “[BOS]” digunakan untuk menandai awal kalimat dan Token “[EOS]” digunakan untuk menandakan akhir kalimat.



Gambar 3.3 Arsitektur Model IEL-TRL

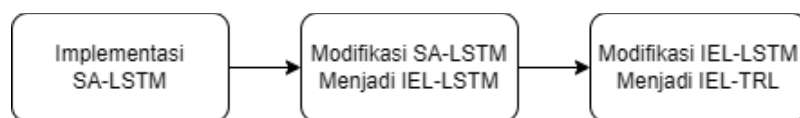
Tahap kedua adalah merancang model yang diajukan. Perancangan model diperoleh dari hasil kajian studi literatur penelitian terdahulu mengenai model yang sesuai untuk menyelesaikan topik penelitian. Penelitian ini mengajukan dua model yang dihipotesiskan mampu menjawab pertanyaan rumusan masalah. Rancangan model pertama dinamakan IEL-TRL dan dapat dilihat pada Gambar 3.3. Model ini didasari penggabungan *encoder* penelitian (Yang dkk., 2021) dengan *decoder* dari (Zhang dkk., 2020). Selanjutnya pada model kedua dinamakan ORG-NACF dan didasari oleh penggabungan *encoder* khususnya *Object Relational Graph*

(ORG) penelitian (Zhang dkk., 2020) dengan *decoder Transformer* dari (Yang dkk., 2021) yang dapat dilihat pada Gambar 3.4.



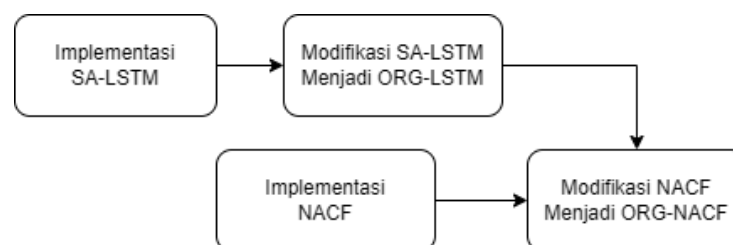
Gambar 3.4 Arsitektur Model ORG-NACF

Tahap ketiga adalah eksekusi skenario pemodelan merujuk pada rancangan model yang telah dibuat. Setiap model memiliki skenario pemodelan yang berbeda. Model IEL-TRL dikembangkan dari model *Soft-Attention LSTM* (SA-LSTM) (Yao dkk., 2015). Model SA-LSTM ini bertindak sebagai *baseline* untuk Model IEL-TRL. Selanjutnya, Model SA-LSTM dikembangkan dengan menambah alur data *motion feature*. Kemudian diimplementasikan metode *Teacher Recommended Learning* (TRL) terhadap *training* model. Metode TRL memanfaatkan *External Language Model* (ELM) untuk membangkitkan *soft target* ketika *training*. ELM yang digunakan adalah DistilBERT yang dilatih dengan *caption* MSRVT. Alur skenario model dapat dilihat pada Gambar 3.5



Gambar 3.5 Skenario pemodelan IEL-TRL

Skenario pemodelan ORG-NACF memiliki empat tahap utama. Model ORG-NACF dikembangkan berdasarkan implementasi penelitian (Zhang dkk., 2020) dengan mengambil modul *Object Relational Graph* (ORG) dan (Yang dkk., 2021) berjudul *Non-Autoregressive Coarse to Fine* (NACF) *Video Captioning*. Model NACF akan bertindak sebagai *baseline* ORG-NACF. *Baseline* ini kemudian dimodifikasi dengan menambahkan modul ORG. Modul ORG perlu dikembangkan terlebih dahulu dari model SA-LSTM sebelum dipasangkan dengan NACF. Hal ini dilakukan untuk menguji keberadaan dan besaran pengaruh dari modul ORG. Modul ORG akan dipasangkan dengan NACF jika memperoleh hasil perkembangan yang representatif. Alur skenario ini dapat dilihat pada Gambar 3.6.



Gambar 3.6 Skenario pemodelan ORG-NACF

### 3.1.5 Pemodelan *Video Captioning*

Pemodelan *video captioning* adalah tahap implementasi rancangan model dan pelaksanaan skenario pemodelan. Berdasarkan skenario pemodelan IEL-TRL, pertama diimplementasikan model *baseline* yaitu SA-LSTM. Selanjutnya, model divalidasi arsitekturnya sesuai penelitian (Yao dkk., 2015). *Training* akan dilakukan terhadap model SA-LSTM jika berhasil divalidasi. Model yang sudah dilakukan *training* kemudian dievaluasi. Unjuk kerja hasil evaluasi tersebut dijadikan dasar pengembangan model IEL-TRL.

Pengembangan tahap pertama adalah Model SA-LSTM dimodifikasi agar dapat menerima *motion feature* menjadi Model IEL-LSTM. Model baru ini kemudian diverifikasi apakah model dapat mengolah *motion feature*. Model akan dilatih jika proses verifikasi telah berhasil. Model baru yang telah selesai dilatih kemudian dievaluasi. Unjuk kerja hasil evaluasi ini dibandingkan dengan unjuk kerja dari model *baseline*. Hal ini dilakukan untuk mengetahui pengaruh dari pengembangan model dan langkah yang perlu diambil selanjutnya.

Modifikasi selanjutnya adalah penambahan metode TRL pada saat *training*. Metode TRL menggunakan ELM untuk membangkitkan *soft target*. ELM ini perlu dilatih terlebih dahulu terhadap *caption* MSR-VTT. ELM yang telah dilakukan *fine tune* digabungkan dengan model IEL-LSTM menjadi IEL-TRL. Model baru ini kemudian dilakukan *training* dan dievaluasi. Unjuk kerja hasil evaluasi ini dibandingkan dengan model *baseline* dan model IEL-LSTM.

Pemodelan ORG-NACF didasari rancangan model dan skenario pemodelan. Skenario pemodel ORG-NACF dimulai dari implementasi model *baseline* NACF. Model ini kemudian divalidasi arsitekturnya berdasarkan penelitian (Yang dkk., 2021). *Training* Model NACF dilakukan jika proses validasi telah berhasil. Model yang sudah dilatih kemudian dievaluasi. Unjuk kerja hasil evaluasi ini dijadikan sebagai *baseline* dalam pengembangan Model NACF.

Modifikasi Model NACF dilakukan dengan menggabungkan modul ORG dengan NACF menjadi Model ORG-NACF. Modul ORG diverifikasi apakah hasil masukan dan keluaran telah sesuai dengan teori dari penelitian (Zhang dkk., 2020). Setelah verifikasi berhasil, model divalidasi terhadap rancangan model. *Training* dilakukan jika validasi model telah selesai. Model ini kemudian dievaluasi dan hasil evaluasi dibandingkan dengan model *baseline*.

### 3.1.6 Perancangan Skenario Eksperimen

Subbab ini akan menjelaskan perancangan skenario eksperimen. Penelitian ini akan melakukan tiga buah skenario di antaranya skenario penggunaan jumlah data, skenario eksperimen arsitektur, dan skenario eksperimen *baseline*. Perancangan skenario ini dilakukan untuk menentukan konfigurasi *hyperparameter* yang cocok untuk setiap eksperimen dan juga dapat dijadikan dasar pengambilan keputusan saat eksperimen.

#### 3.1.6.1 Skenario Penggunaan Jumlah Data

Dataset yang digunakan pada penelitian ini merupakan dataset MSR-VTT yang berjumlah sebanyak 10.000 data video. Skenario ini membagi data menjadi tiga varian di antaranya varian 100 data, 1.000 data, dan 10.000 data. Pengambilan data video dilakukan secara acak dari dataset utama MSR-VTT. Secara rinci skenario ini dapat dilihat pada Tabel 3.1. Skenario bertujuan untuk mengetahui

kemampuan setiap model terhadap variasi jumlah data dan sebagai dasar konfigurasi *hyperparameter* yang digunakan untuk tahap *training* selanjutnya.

Tabel 3.1  
Rincian Skenario Penggunaan Jumlah Data

No.	Jumlah Variasi Data Video	Rasio Pembagian Himpunan Data
1	100 Data Video	85% : 5% : 10%
2	1.000 Data Video	85% : 5% : 10%
3	10.000 Data Video	65% : 5% : 30%

### 3.1.6.2 Skenario Eksperimen Arsitektur

Skenario eksperimen arsitektur memuat implementasi, pelatihan, dan pengembangan model yang diajukan. Model yang diajukan adalah IEL-TRL dan ORG-NACF. Arsitektur model tersebut dapat dilihat pada Gambar 3.3 dan 3.4. Skenario ini bertujuan untuk mengatur *hyperparameter* yang agar diperoleh model yang optimal. Skenario ini digabungkan dengan skenario penggunaan jumlah data.

### 3.1.6.3 Skenario Eksperimen *Baseline*

Skenario eksperimen *baseline* membandingkan hasil model yang diperoleh dari skenario eksperimen arsitektur dengan model *baseline*. Model *baseline* tersebut adalah ORG-TRL dan NACF. Unjuk kerja dari model *baseline* akan menjadi dasar pembandingan terhadap model yang diajukan. Hasil evaluasi ini bertujuan untuk mengetahui apakah model yang diajukan memperoleh unjuk kerja yang lebih baik dengan waktu *inference* yang lebih cepat atau sebaliknya.

### 3.1.7 Eksperimen

Hasil rancangan metode dan rancangan skenario eksperimen dijadikan sebagai acuan ketika eksperimen dilakukan. Tahap eksperimen memuat praproses data, implementasi model, pengembangan model, perbaikan model, dan evaluasi model. Hasil yang diperoleh dari tahap ini adalah model yang sudah dilatih, hasil unjuk kerja model, dan hasil pembangkitan *caption* dari model.

### 3.1.8 Analisis dan Evaluasi

Hasil yang diperoleh dari tahap eksperimen kemudian dianalisis dan dievaluasi. Analisis dilakukan terhadap hasil unjuk kerja, hasil pembangkitan



*caption*, dan model yang diajukan. Evaluasi dilakukan terhadap unjuk kerja dan hasil pembangkitan *caption*. Evaluasi kuantitatif dilakukan dengan menggunakan metrik evaluasi otomatis seperti BLEU, CIDEr, ROUGE-L, dan METEOR. Evaluasi kualitatif dilakukan dengan menganalisis hasil pembangkitan *caption* dari model yang diajukan.

### 3.1.9 Penarikan Kesimpulan

Penarikan kesimpulan dilakukan ketika semua tahapan penelitian telah berhasil dilakukan. Selain itu, peneliti perlu memberikan saran agar dapat bermanfaat untuk penelitian selanjutnya.

## 3.2 Alat dan Bahan Penelitian

Bagian ini akan menjelaskan alat dan bahan yang digunakan selama penelitian berlangsung.

### 3.2.1 Alat Penelitian

Penelitian ini menggunakan alat-alat meliputi perangkat keras dan perangkat lunak dalam upaya untuk mencapai tujuan penelitian. Berikut adalah daftar perangkat dan data yang akan digunakan dalam jalannya penelitian:

1. Perangkat keras (Hardware) yang digunakan yaitu sebuah laptop dengan spesifikasi:
  - a. *Processor* AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz
  - b. *Random Access Memory* (RAM) 12 GB
  - c. AMD Radeon(TM) Vega 8 *Graphics*
2. Perangkat lunak (Software) yang digunakan:
  - a. Jupyter Notebook
  - b. Google Collab
  - c. Kaggle
  - d. Papers Gradient
  - e. Anaconda Command Prompt
  - f. Microsoft Edge
  - g. GitHub
  - h. Windows Command Prompt

### 3.2.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah dataset *Microsoft Research Video to Text* (MSR-VTT). Bahan ini terdiri dari pasangan video dengan beberapa *caption* yang beragam.