

# 도메인 특화 언어 모델을 이용한 항공 안전 보고서 내 원인 요인 식별

## Identification of Causal Factors in Aviation Safety Reports using Domain-specific Language Model

이동건, 김태형  
(DongGeon Lee, TaeHyung Kim)

지도교수: 최원익 (Wonik Choi)

**요약:** 본 논문은 딥러닝을 통한 항공 안전 보고서 내 원인 요인 식별 시스템을 제안한다. 먼저 한국어 항공 안전 도메인 데이터셋을 구축하고, Domain Adaptive Pre-training을 통해 항공 안전 도메인에 특화된 KoAirBERT 모델을 개발한다. 이후 지도학습 및 비지도 방식으로 KoAirBERT를 활용하여 원인 요인을 식별하고, 두 방식의 결과를 비교한다. 가장 우수한 성능을 보인 Fine-tuned-KoAirBERT는 0.8103의 F1-score를 달성하였으며, 식별한 원인 요인과 비(非)원인 요인 간의 차이를 t-SNE를 통해 시각화하여 확인한다.

**Abstract:** In this paper, we propose a deep learning system for identifying causal factors in aviation safety reports. First, we build a Korean aviation safety domain dataset, and develop a KoAirBERT model specialized for the aviation safety domain using Domain Adaptive Pre-training. Then, we identify causal factors using KoAirBERT in a supervised and unsupervised method, compare the results of both methods. The best performing method, Fine-tuned-KoAirBERT, achieved an F1 score of 0.8103, and we visualize the difference between the causal and non-causal factors identified by the model using t-SNE.

**Keywords:** Aviation Safety Reports, Identification of Causal Factors, BERT, Domain Adaptive Pre-training, Keyphrase Extraction

### 1. 서론

항공 운송 기술 발달로 국내 항공산업은 비약적인 발전을 이루었지만, 항공 안전사고 및 장애 발생률 역시 증가하고 있다. 이에 따라, 안전사고를 선제적으로 예방할 수 있도록 항공 데이터 수집 및 분석에 대한 필요성이 점차 대두되고 있다 [1]. 이에 한국교통안전공단은 항공 안전을 저해하거나, 저해할 우려가 있는 사건이나 상태 또는 상황을 자율 신고를 통해 보고받는 제도인 항공 안전 자율보고제도를 운영 중이다 [2]. 또한 항공 사고 예방을 위한 연구의 기초자료로 활용할 수 있도록, 접수된 항공 안전 자율보고서와 해외의 항공 안전 보고 사례들을 종합적으로 수집하여 GYRO (항공 안전 자율보고 정보지)를 분기 별로 발간하고 있다 [3,4].

항공 안전과 관련된 문제점 개선 및 사고 예방을 위해선, 접수된 보고서를 통해 사후적으로 사고 유발 원인을 파악함으로써 사고를 사전에 예측하는 것이 필요하다 [1]. 이러한 사고 원인 분석은 주로 항공 안전 전문가들에 의해 수행되고 있으며, 인공지능을 통한 원인 분석을 도입하기 위한 연구들이 이뤄지고 있다 [5-7]. [7]에서는 언어 모델을 이용하여 항공 안전 데이터를 항공 안전 데이터 표준 분류체계에 따라 정형적으로 분류하였다. 하지만 이는 데이터 자체를 분류한 것이므로, 보고서 내에서 구체적인 원인을 파악하기 위해선 전문가가 다시 보고서를 분석해야 한다는 한계가 있다. 이에 본 논문에서는 선제적으로 항공

안전사고에 대응하고, 보고서의 원인 분석에 소요되는 인적 비용을 줄이고자, 항공 안전 보고서에 기술된 사고 발생 원인 요인을 식별하는 딥러닝 기반의 시스템을 제안한다. 이 논문의 기여는 다음과 같이 요약된다.

- 1) 도메인 특화 언어 모델 개발을 위해 1.9만 건의 데이터를 수집하여 한국어 항공 안전 도메인 데이터셋을 구축한다.
- 2) 항공 안전 도메인에 특화된 언어 모델인 KoAirBERT 모델을 제안한다. KoAirBERT 모델은 한국어 Bidirectional Encoder Representations from Transformers (BERT) 모델인 KLUE-BERT-base [8]를 기반으로 Domain Adaptive Pre-training (DAPT) [9]을 수행한 모델로, 항공 안전 분야와 관련된 과업에서 기존 KLUE-BERT-base보다 높은 성능을 보인다.
- 3) 보고서 내 사고 원인 요인 식별을 위해, 지도학습 및 비지도학습 기반 실험을 진행한다. 지도학습 기반의 실험은 KoAirBERT 모델을 미세조정 (Fine-Tuning) 하여 원인을 서술하는 키프레이즈 (Keyphrase)를 추출하고, 비지도학습 기반의 실험은 KeyBERT [10] 모델을 응용하여 수행한다. 두 실험에서 원인 요인을 식별한 결과를 정량적 및 정성적으로 비교하며 지도학습을 사용한 원인 요인 식별이 더 적합함을 보인다.
- 4) 식별한 원인 요인과 비(非)원인 요인 간의 차이를 차원 축소 알고리즘인 t-SNE와 군집화 알고리즘인 K-means를 통해 시각화하여 비교한다.

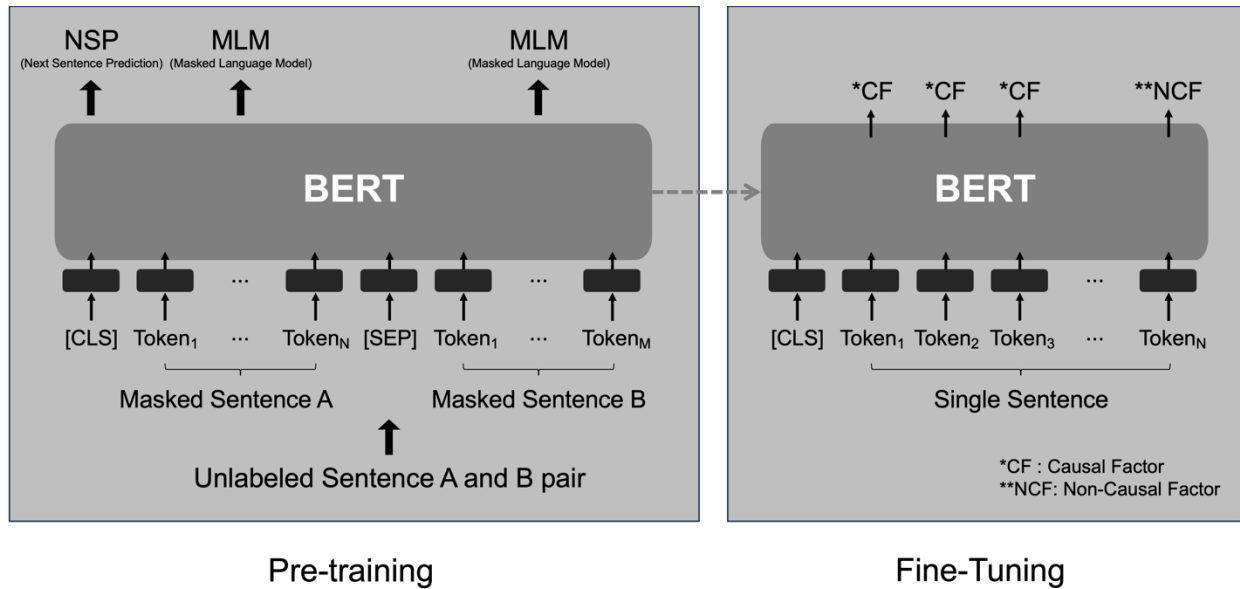


그림 1. BERT의 전반적인 사전학습 (Pre-training) 및 미세조정 (Fine-Tuning) 과정

## II. 관련 연구

### 1. BERT

Bidirectional Encoder Representations from Transformers (BERT)는 구글이 발표한 사전학습된 (Pre-trained) 언어모델로 자연어 처리 분야에서 뛰어난 성과를 보이고 있다 [11]. 이전의 많은 모델들은 문장을 한 방향으로만 처리했지만 [12, 13], BERT는 양방향 Transformer 인코더를 사용함으로써 문장의 양쪽으로 동시에 처리하여 문맥을 더욱 정확하게 파악할 수 있다.

BERT는 그림 1처럼 사전학습 (Pre-training)과 미세조정 (Fine-Tuning)의 두 단계로 모델을 훈련한다. 사전학습은 레이블링이 되지 않은 대량의 텍스트 데이터로 모델을 학습하는 과정이며, 세부적으로 Masked Language Model (MLM)과 Next Sentence Prediction (NSP)의 두 가지 학습 방법을 통해 이루어진다. MLM은 문장 내의 일부 단어를 무작위로 가리고, 이 가려진 단어를 예측하는 방식으로 학습하는 방법이다. NSP는 두 문장이 주어졌을 때, 두 번째 문장이 첫 번째 문장의 바로 다음에 오는 문장인지에 대한 여부를 예측하며, 이를 통해 BERT는 두 문장 사이의 관계를 이해할 수 있다. 미세조정은 사전학습된 모델을 특정 작업에 맞도록 추가적인 훈련을 통해 조정하는 단계이며, 사전학습된 BERT 모델에 추가적인 계층을 붙여 특정 작업에 최적화된 모델을 만든다. 이는 작업에 따라 개체명 인식 (Named Entity Recognition), 토큰 분류 (Token Classification) 등 다양한 형태로 이루어질 수 있다. 일반적으로 미세조정 단계에서는 적은 양의 데이터로도 높은 성능을 보이는데, 이는 BERT가 사전학습 단계에서 축적한 언어 이해 능력을 활용하기 때문이다.

### 2. KLUE-BERT-base

Korean Language Understanding Evaluation (KLUE)는 한국어

자연어 처리 모델의 성능을 평가하고 개선하기 위한 벤치마크이며, 문장 분류, 관계 추출, 개체명 인식, 질의응답 등 다양한 문제 영역을 포함하고 있다 [8]. KLUE-BERT-base는 원래의 BERT와 동일한 아키텍처를 사용하며, KLUE 벤치마크를 학습하며 한국어 텍스트에 최적화된 학습을 수행한 모델이다 [8].

### 3. Domain Adaptive Pre-training (DAPT)

Post-training은 미세조정의 성능을 개선하기 위해 고안된 방법으로, 사전학습 후 다른 데이터셋으로 추가적인 학습을 수행하며 모델을 다른 데이터셋에 적응시키는 (adapted) 방법이다 [14, 15]. Post-training 기법의 하나인 Domain Adaptive Pre-training (DAPT)은 미세조정에 사용하고자 하는 도메인에 특화된 말뭉치를 Post-training의 학습 데이터셋으로서 사용하는 방법이다 [9]. 이전 연구 [16, 17]에서는 BERT를 이용한 Post-training을 제안하였으며, 도메인에 특화된 말뭉치로 Post-training을 수행했을 때 일반적인 말뭉치에 나타나지 않는 표현들과 단어들을 훈련할 수 있다는 것을 검증하였다.

### 4. KeyBERT

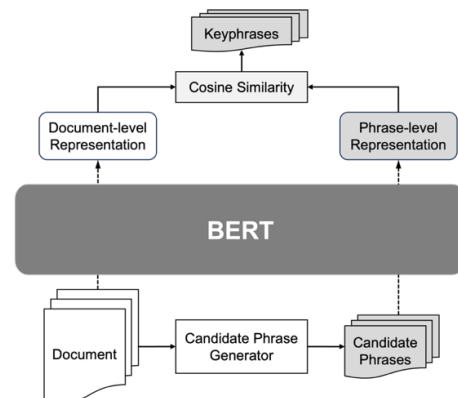


그림 2. KeyBERT 시스템의 키프레이즈 추출 과정

EmbedRank는 문서와 후보 구문 (Candidate Phrase)의 임베딩 벡터들을 구하고, 임베딩 벡터들 간의 코사인 유사도를 구함으로써 키프레이즈 (Keyphrase)를 추출하는 비지도학습 기반의 모델이다 [18]. KeyBERT는 EmbedRank를 응용한 시스템으로, BERT 기반 모델을 사용하여 문서와 후보 구문의 임베딩 벡터를 구하여 키프레이즈를 추출하는 시스템이다 [10]. 이를 통해 비지도학습의 특징은 유지하며, EmbedRank보다 문서의 전체적인 문맥을 더욱 정확하게 파악하여 반영할 수 있다는 장점을 가진다. 그림 2는 KeyBERT 시스템이 입력 문서의 키프레이즈 추출하는 과정을 나타낸다.

### III. 연구 방법

3장에서는 본 논문에서 수행된 연구 방법을 그림 3과 같이 3단계로 나누어 서술한다.

### 1. 데이터셋 수집 및 구축

본 연구에서 구축한 한국어 항공 안전 도메인 데이터셋은 항공우주산업융합원, 항공철도사고조사위원회, 한국교통안전공단 및 웹 크롤링을 통해 수집하였으며, 각 데이터의 정보는 표 1에 나타내었다. 데이터셋 중 안전 장애 보고 및 고장 보고 데이터는 ‘항공안전법 제2조 10의4 가목’에서 정의하는 항공 안전 데이터이며, 사고·준사고 보고서 데이터는 [19]에서 제공하는 2005년부터 2019년 사이의 보고서 일부를 사용한 것이다. 항공 안전문화 지표 분석 데이터는 GYRO 제1호부터 GYRO 제206호까지의 항공 안전 자율보고서 [3] 중 일부를 항공 안전문화 지표에 따라 분석한 분석 내용이고, GYRO 본문 증강 데이터는 GYRO를 통해 수집한 항공 안전 자율보고서를 역 번역 (Back Translation) 방식으로 증강한 것이다. 항공 용어사전 [20] 및 항공위키 [21] 데이터는 웹 크롤링을 통해 수집하였다.

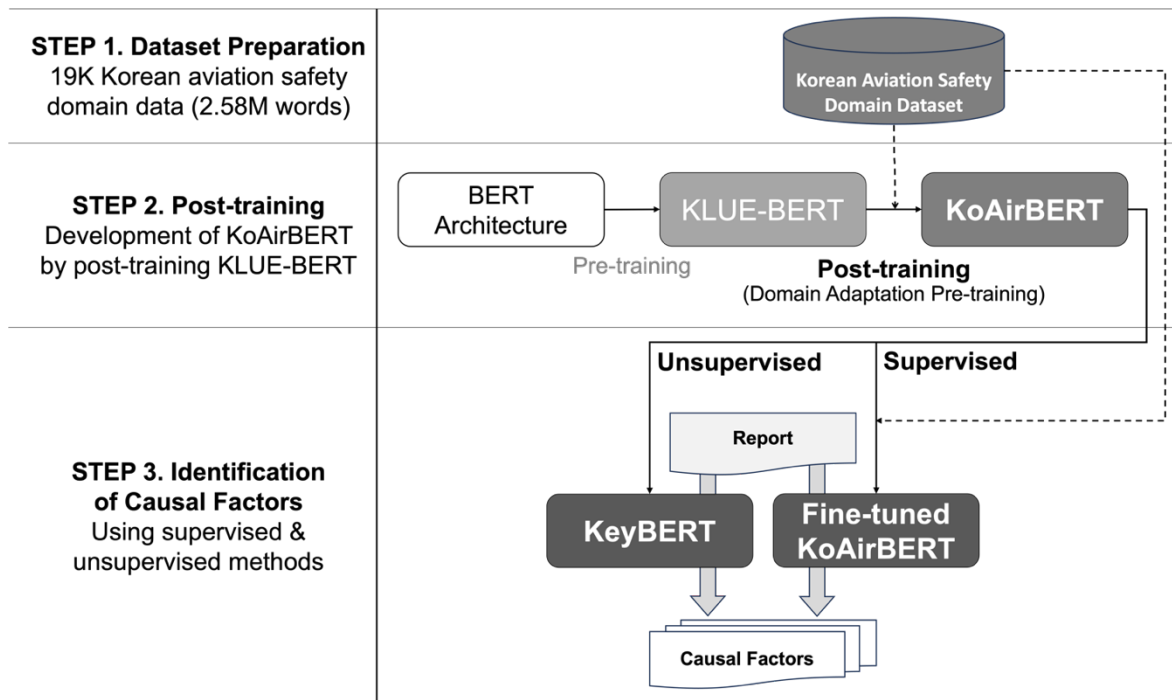


그림 3. 연구 절차 단계

표 1. KoAirBERT 학습에 사용된 한국어 항공 안전 도메인 데이터셋의 정보

데이터명	데이터 건 수	문장 수	단어 수
안전 장애 보고 데이터	684	10,018	69,472
고장 보고 데이터	1,771	10,480	85,165
사고·준사고 보고서 [19]	54	1,850	33,935
항공 안전문화 지표 분석 데이터 [3]	1,055	3,652	244,032
GYRO 항공 안전 자율보고서 [3] - 본문 증강 데이터	6,776	66,848	1,214,061
항공정보포털시스템 항공 용어사전 [20]	4,961	15,312	167,295
항공위키 [21]	4,314	38,927	766,214
합계	19,615	147,087	2,580,174

## 2. KoAirBERT 개발

항공 안전 분야에 특화되지 않은 일반적인 언어 모델은 항공 도메인 데이터에 자주 등장하는 특수한 고유명사나 전문용어로 학습되지 않았기 때문에 좋은 성능을 기대하기 힘들다. 따라서 본 연구는 항공 안전 보고서에서의 원인 요인을 식별에 앞서, 항공 안전 분야에 특화된 KoAirBERT 모델을 개발한다. 이를 위해 한국어 데이터로 사전학습된 KLUE-BERT-base [8] 모델을 기본 모델로 사용하였으며, KLUE-BERT에 자체 구축한 한국어 항공 안전 도메인 데이터셋을 활용해 Domain Adaptive Pre-training (DAPT) [9] 방식의 Post-training을 수행하였다. Post-training 시에는 MLM과 NSP 방법을 모두 사용하여 토큰들 및 문장들 사이의 상관관계를 해석할 수 있도록 모델을 학습하였으며, 학습 시 설정한 하이퍼파라미터는 표 2에 나타내었다.

표 2. Post-training에 사용된 하이퍼파라미터 값

하이퍼파라미터	값
Initial Learning Rate	5e-5
Weight Decay	0.01
Max Sequence Length	512
Batch Size	8
Epochs	3

## 3. 원인 요인 식별

본 연구에서는 개발한 KoAirBERT 모델을 지도학습 방식과 비지도 방식으로 활용하여 원인 요인을 식별한다.

### 1) 지도학습 방식 (미세조정)

지도학습 방식의 원인 요인 식별을 위해, 언어 모델을 토큰 분류 (Token-Classification) 방식으로 미세조정 (Fine-tuning)한다. 토큰 분류란 입력 보고서를 토큰화 (tokenization)한 후, 토큰 단위로 원인 요인과 비(非)원인 요인을 분류하는 것을 말한다.

### 2) 비지도 방식 (KeyBERT)

비지도 방식의 원인 요인 식별을 위해서는 KeyBERT 시스템을 통해 키프레이즈 (Keyphrase)를 추출한다. 키프레이즈 추출 시에 KoAirBERT를 활용하여 문서와 후보 구문의 임베딩을 구하며, 키프레이즈의 추출 결과를 다양화하기 위한 방법으로 Max Sum Similarity (MSS) 알고리즘을 사용한다. MSS 알고리즘이란 문서와 가장 유사성이 높은 키프레이즈들을 넉넉히 추출한 후, 추출한 키프레이즈 집합에서 서로 가장 유사성이 낮은 키프레이즈들을 Top-K개만큼 뽑는 방법이다. MSS 알고리즘을 사용하여 키프레이즈를 추출하면, 주어진 문서와 가장 관련이 높은 키프레이즈를 찾을 수 있으면서도 키프레이즈 간 유사성을 낮추어 다양성을 높일 수 있다는 장점을 가진다.

## IV. 실험 및 결과

### 1. 실험 및 평가 대상

본 절에서는 원인 요인 식별 과업을 평가하기 위해 지도학습 방식과 비지도 방식을 설명한다.

#### 1) 지도학습 방식

지도학습 방식의 원인 요인 식별 실험을 위해 언어 모델을 미세조정 (Fine-tuning)하였으며, 제안한 항공 안전 도메인 특화 모델인 KoAirBERT 모델과, 베이스라인 모델로서 한국어 BERT 모델인 KLUE-BERT-base [8] 모델을 사용하여 비교한다. 더불어, 베이스라인 모델보다 발전된 KLUE-RoBERTa-base [8] 모델을 함께 평가하였다.

각 언어 모델들의 미세조정을 위해 GYRO (항공 안전 자율보고 정보지) [3]를 통해 수집한 항공 안전 자율보고서 중 사고 원인이 명확히 기술된 1,093건을 사용하였다. 보고서마다 사고 원인을 직접 라벨링하였으며, 라벨링을 위해 GYRO에 기술된 각국의 항공 안전 보고제도 당국의 분석 결과를 참고하였다. 사고 원인이 라벨링 된 항공 안전 자율보고서 1,093건을 Train Set (80%, 874건)과 Test Set (20%, 219건)으로 분리한 후, Train Set을 사용하여 미세조정을 수행하였다.

#### 2) 비지도 방식

비지도 방식으로는 KoAirBERT를 활용한 KeyBERT [10] 시스템을 사용하였다. 원인 요인을 식별하기 위해, 입력 보고서로부터  $n$ -gram ( $n = 3, 4, \dots, 8$ )을 추출하여 후보 구문을 생성한 후, KeyBERT를 활용하여 30개의 키프레이즈들 (Keyphrases)을 추출하였으며, 그 중 Max Sum Similarity 알고리즘을 통하여 상위 5개의 키프레이즈들을 추출하였다.

## 2. 실험 및 평가 결과

### 1) 정량적 평가

모델의 전반적인 성능을 정량적으로 평가하기 위해, 토큰 단위로 원인 요인을 분류한 결과의 Micro-average F1-score를 측정하였다. 표 3은 모델 별 성능을 나타낸다. 본 논문에서 제안하는 지도학습 기반의 Fine-tuned-KoAirBERT 모델은 F1-score 0.8103을 달성하며, 베이스라인 모델보다 좋은 성능을 보인다. 한편, 비지도 기반의 KeyBERT-KoAirBERT 모델은 F1-score 0.7588을 달성하며, 지도학습 기반 모델들 대비 낮은 성능을 보인다.

표 3. 모델 별 성능 측정 결과

Model		F1-score
Supervised	KLUE-BERT-base [8]	0.7985
	KLUE-RoBERTa-base [8]	0.8056
	<b>Fine-tuned-KoAirBERT</b>	<b>0.8103</b>
Unsupervised	KeyBERT [10]-KoAirBERT	0.7588

표 4. 모델 별 원인 식별 결과  
(입력 보고서의 짧은 글자는 정답 원인을 나타냄.)

입력 보고서	이륙 활주 중 No.2 타이어가 파열되었으며, 이로 인해 항공기 동체 하부에 약간의 손상을 입었다. 이륙 후 연료를 Dump 시키고 인천공항에 비상착륙을 시도하였고 착륙은 잘 하였다. 이 과정에서 조종석 내 승무원 간의 CRM 및 객실 승무원들과 Company와도 상호 협조가 매우 잘 되었다.	
Model	Fine-tuned-KoAirBERT	KeyBERT [10]-KoAirBERT
원인 식별 결과	1. 타이어가 파열되었으며, 이로 인해 항공기 동체	1. 약간의 손상을 입었다. 이륙 후 연료를 Dump 2. 시키고 인천공항에 비상착륙을 시도하였고 착륙은 잘 하였다. 3. 상호 협조가 매우 잘 되었다.

## 2) 정성적 평가

본 항에서는 지도학습 기반 모델과 비지도 기반 모델의 성능 차이를 정성적으로 평가하기 위해, Fine-tuned-KoAirBERT 모델과 KeyBERT-KoAirBERT 모델이 동일한 항공 안전 자율보고서에 대해 원인을 식별한 결과를 비교한다. 결과는 표 4에 나타내었다. 표 4에서 제시한 입력 보고서의 정답 원인과 가장 근접하게 원인을 식별한 모델은 지도학습 기반의 Fine-tuned-KoAirBERT이며, 비지도 기반의 KeyBERT-KoAirBERT 모델은 원인보단 결과를 위주로 식별한 것을 확인할 수 있다.

## 3. 임베딩 벡터 시각화 및 분석

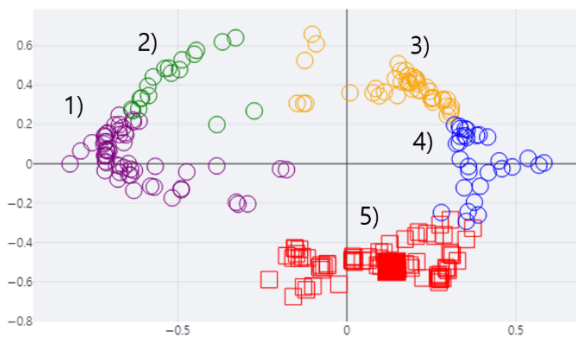


그림 4. t-SNE 및 K-means 알고리즘을 통한 보고서 속 구문 시각화 및 군집화 결과 (원인 구문이 포함된 군집은 사각형 형태의 점, 그 이외의 군집은 원 형태의 점으로 표시함. 원인 구문을 나타내는 점은 꼭 찬 사각형, 원인 구문이 아닌 군집 구성원의 경우는 속이 빈 사각형으로 표시함.)

그림 4는 Fine-tuned-KoAirBERT 모델이 식별한 원인 요인 및 입력 보고서로부터 추출한 후보 구문의 임베딩 벡터 값을 차원 축소 알고리즘인 t-SNE를 통해 2차원 형태의 산점도 형

표 5. 군집 별 주요 구문

군집 번호	주요 구문
1	승무원 간의 CRM
	조종석 내 승무원 간의
2	Company와도 상호 협조가
	객실 승무원들과 Company와도
	협조가 매우 잘 되었다.
3	인천공항에 비상착륙을 시도하였고
	이 과정에서 조종석 내
4	시도하였고 착륙은 잘 하였다.
	연료를 Dump 시키고 인천공항에
5	이륙 후 연료를 Dump
	타이어가 파열되었으며,
	이로 인해 항공기 동체
	항공기 동체 하부에 약간의 손상을
	약간의 손상을 입었다.

태로 나타낸 후, 군집화 알고리즘인 K-means를 통해 5개의 군집으로 군집화한 결과이다. 후보 구문은 입력 보고서로부터 n-gram ( $n=3,4,\dots,8$ )을 추출하여 생성하였다. 표 5는 그림 4에서 나타낸 군집화 결과에서 주요 구문을 군집 별로 정리한 것이다.

그림 4와 표 5를 통해, 5개의 군집이 각각 다른 글 흐름을 의미하며, KoAirBERT가 찾아낸 사고 원인은 5번 군집에 속하는 것을 확인할 수 있다. 표 5에서 원인이 속한 5번 군집은 타이어 파열로 인해 항공기에 손상을 입었다는 부분을 의미하며, 그와 반대로 나머지 군집들은 각각 사고 이후 승무원의 대처, Company와의 협조, 인천공항으로 비상착륙, 비상착륙 전 연료 Dump와 같이 사고의 원인과는 거리가 먼 구문의 군집됨을 확인할 수 있다.

## V. 결론

본 연구에서는 보고서 내 원인 요인 식별 과업을 위해, 1.9만 건의 한국어 항공 안전 도메인 데이터셋을 구축하고 항공 안전 분야 특화 언어 모델인 KoAirBERT를 개발하였다. 또한 원인 요인 식별 과업에 대해서, 토큰 분류 (Token-Classification) 방식으로 미세조정 (Fine-tuning)한 Fine-tuned-KoAirBERT 모델을 통해 0.8103의 F1-score을 달성하였다. 또한, 실험을 통해 지도학습 방식 모델의 원인 식별 결과는 사고 원인을 중심으로 예측되는 것을, 비지도 방식의 모델의 식별 결과는 사고 결과를 중심으로 예측되는 것을 보였다.

항공 안전 분야에서 사고 원인 요인을 빠르고 정확하게 식별하는 것은 항공 안전사고의 예방과 항공산업의 안전성 향상을 기대할 수 있다. 개발한 원인 요인 식별 시스템이 실무에 적용된다면, 항공 안전 보고서 분석에 소요되는 시간적·인적 비용을 줄일 수 있을 것으로 기대한다.

## 참고문헌

- [1] 김지연, and 박노삼, “항공안전데이터 및 분석 동향,” *전자통신동향분석*, vol. 36, no. 6, pp. 55-66, 12, 2021.
- [2] 한국교통안전공단. "항공안전 자율보고제도," 2023; <https://www.airsafety.or.kr>.
- [3] 한국교통안전공단, *GYRO*, no. 1-206, 2000-2021.
- [4] 김광중, *항공안전자율보고 백서*, 한국교통안전공단 철도항공안전본부 항공안전처, 경상북도, 2021.
- [5] 한국과학기술원, *빅데이터 기반 항공안전관리 기술 개발 기획 최종보고서*, 17RDPP-C136026-01, 국토교통부, 2018.
- [6] NASA TechPort. "DAAS: Data Analytics for Assurance of Safety, Phase II," 12, 2023; <https://techport.nasa.gov/view/93435>.
- [7] 양성훈 *et al.*, “자연어처리 알고리즘을 이용한 위험기반 항공안전데이터 자동분류 방안 연구,” *한국항공학회논문지*, vol. 26, no. 6, pp. 528-535, 2022.
- [8] S. Park *et al.*, "KLUE: Korean Language Understanding Evaluation," *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1 (NeurIPS Datasets and Benchmarks 2021)*, 2021.
- [9] S. Gururangan *et al.*, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 8342-8360, 2020.
- [10] M. Grootendorst, "KeyBERT: Minimal keyword extraction with BERT," Zenodo, 2020.
- [11] J. Devlin *et al.*, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. pp. 4171-4186, 2019.
- [12] S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [13] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1724-1734, 2014.
- [14] M.-T. Luong, and C. Manning, "Stanford Neural Machine Translation Systems for Spoken Language Domains," *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign*. pp. 76-79, 2015.
- [15] H. Xu *et al.*, "BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. pp. 2324-2335, 2019.
- [16] T. Whang *et al.*, "An Effective Domain Adaptive Post-Training Method for BERT in Response Selection," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. pp. 1585-1589, 2020.
- [17] 김균엽 *et al.*, "Post-training 을 이용한 한국어 전이 학습 모델에서의 성능 향상," *한국정보과학회 2021 한국컴퓨터종합학술대회 논문집*. pp. 301-303, 2021.
- [18] K. Bennani-Smires *et al.*, "Simple Unsupervised Keyphrase Extraction using Sentence Embeddings," *Proceedings of the 22nd Conference on Computational Natural Language Learning*. pp. 221-229, 2018.
- [19] 항공철도사고조사위원회. "항공사고조사보고서," [https://araib.molit.go.kr/USR/airboard0201/m\\_34497/1st.jsp](https://araib.molit.go.kr/USR/airboard0201/m_34497/1st.jsp).
- [20] 항공정보포털시스템. "항공정보포털 전자도서관 항공용어사전," 2023; <https://www.airportal.go.kr/knowledge/library/KdMain01.jsp>.
- [21] "항공위키," 11, 2023; <https://airtravelinfo.kr/wiki/>.



**이 동 권**

2018년 ~ 현재 인하대학교 정보통신공학과 학사과정 재학중.

2024년 2월 졸업예정.

관심분야는 인공지능, 자연어 처리.



**김 태 형**

2018년 ~ 현재 인하대학교 정보통신공학과 학사과정 재학중.

2024년 2월 졸업예정.

관심분야는 인프라, 데이터베이스, 운영체제, 데이터 시각화.