

# LG Aimers 온라인 채널 제품 판매량 예측 AI 온라인 해커톤

김찬호 변유정 부도현 윤채영

# SME AIR

# 목차

- Introduction

- Introduction
- EDA

- Data Preprocess

- Feature Engineering

- Model

- Model Architecture
- Train Method
- Validation Method

- Result&Application

- Q&A



# Introduction

# Introduction



“온라인 채널 제품 판매량 예측 AI”

#시계열

#판매량(수요) 예측

# AI

## DATASET

### train.csv

ID: 28,894개  
대분류: 5개  
중분류: 11개  
소분류: 53개  
브랜드: 2,895개  
쇼핑몰: 10개  
제품: 13,896개  
2022-01-01  
~2023-04-24  
일일 판매량 데이터

### sales.csv

ID: 28,894개  
대분류: 5개  
중분류: 11개  
소분류: 53개  
브랜드: 2,895개  
쇼핑몰: 10개  
제품: 13,896개  
2022-01-01  
~2023-04-24  
일일 판매액 데이터

### brand\_keyword\_cnt.csv(메타)

브랜드: 3,822개  
결측치: 41개  
2022-01-01~2023-04-24  
브랜드 연관키워드 정규화 언급량

### product\_info.csv(메타)

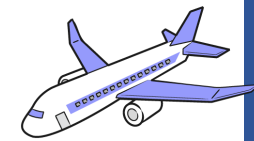
제품: 12,778개  
제품 정보 데이터



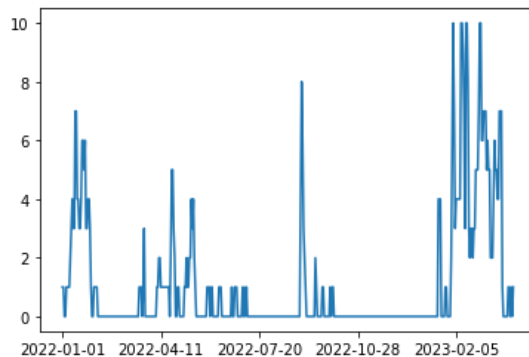
**EDA**



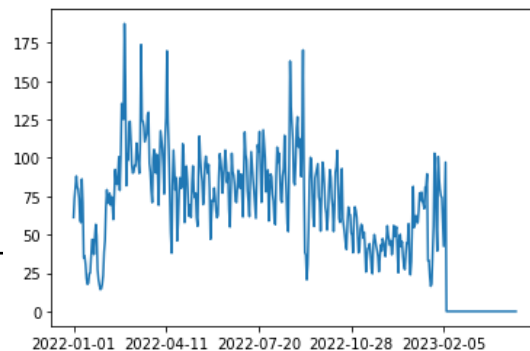
# EDA-1



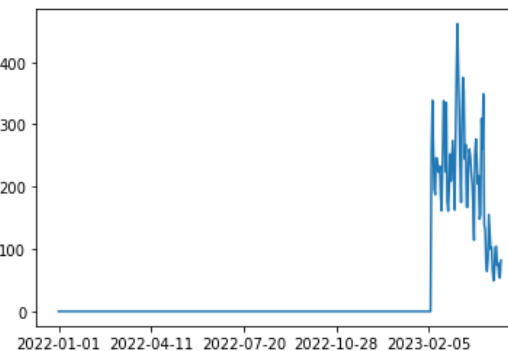
train.csv



판매량 = 0 다수 식별

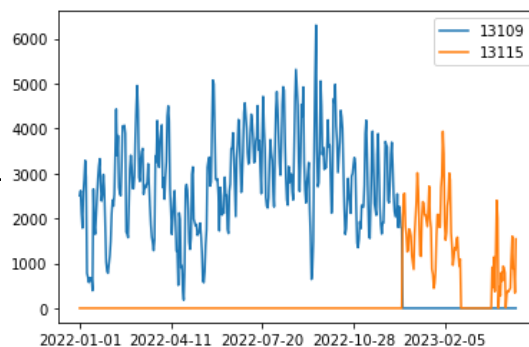


4번 쇼핑물 평균 판매량



10번 쇼핑물 평균 판매량

4번, 10번 쇼핑물  
2023/02/08기준  
판매량 이전 식별

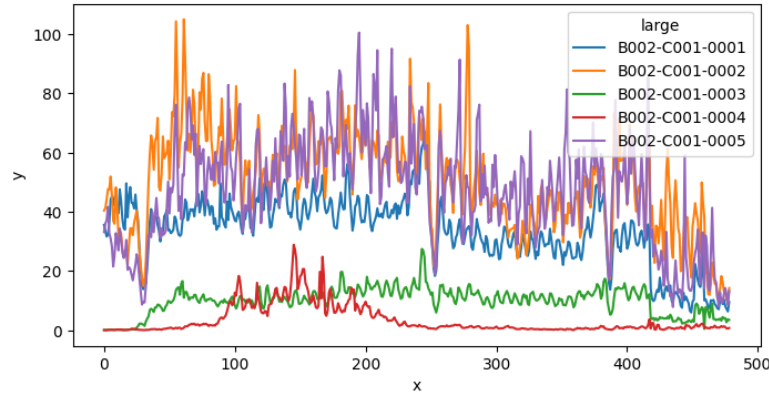


동일 {제품,쇼핑물} 31쌍  
판매량 이전 식별

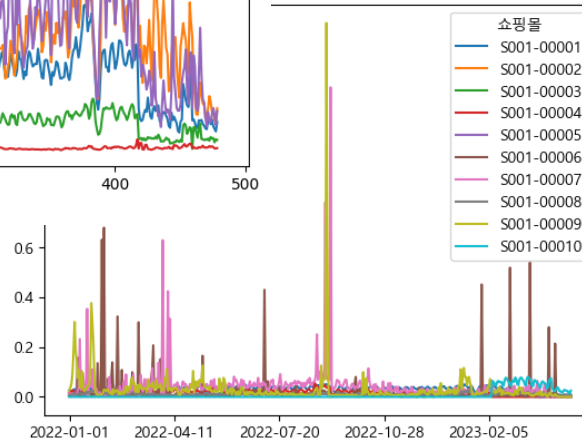
# EDA-2



train.csv



대분류별 판매량 평균 그래프

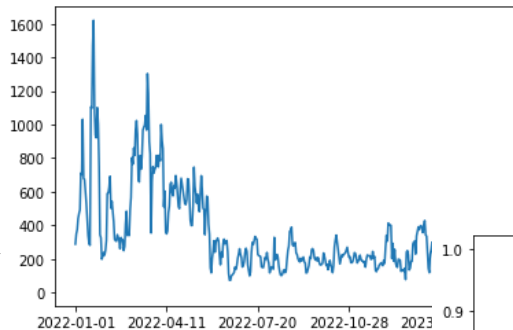


쇼핑몰별 판매량 평균 그래프

대/중/소분류, 쇼핑몰 별  
판매량 추이 상이 식별

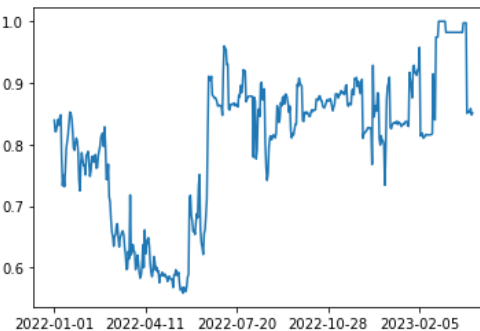
train.csv

sales.csv



판매량 변동

개당 가격 변동



개당 가격 변동에 따른  
판매량 변동 식별

# EDA-3

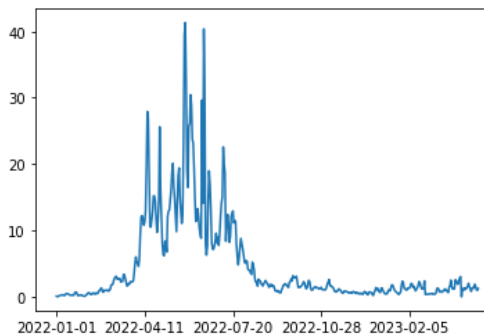


product\_info.csv

대분류/중분류/소분류  
WORDCLOUD

분류 구분 및  
트렌드, 추세 식별

train.csv



Ex) 대분류 4의 중분류 9는  
선크림이라서 여름에 잘 팔렸겠구나?





# Feature Engineering



## 판매량

## NULL\_index

## Price\_index

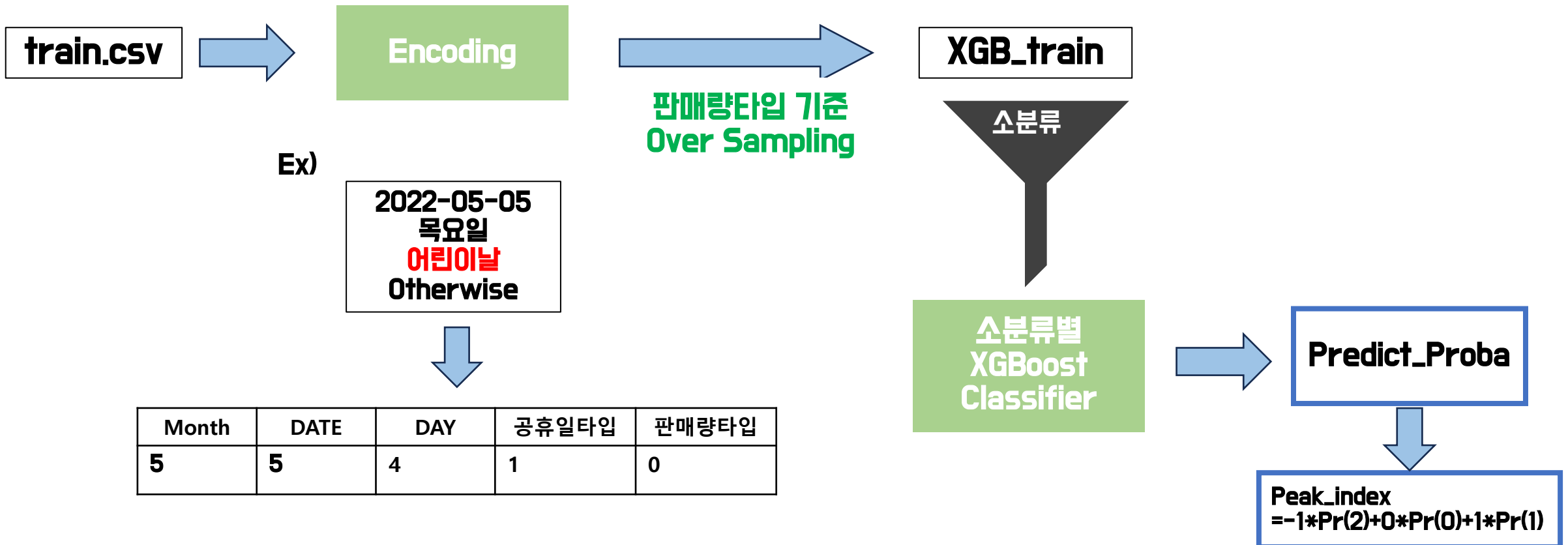
## 대/중/소분류& 쇼핑몰 MDS

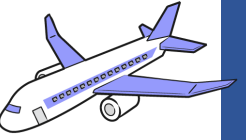
## HOT\_index



# Feature Engineering-2

Feature	Month(X1)	DATE(X2)	DAY(X3)	공휴일타입(X4)	판매량타입(Y)
Content	월	일자	월요일부터 일요일까지 1부터 7까지로 인코딩	설날, 추석 연휴 -> 2 설날, 추석 제외 공휴일-> 1 Otherwise-> 0	제품별 판매량 하위 20% 기간 -> 2 제품별 판매량 상위 20% 기간 -> 1 Otherwise -> 0





# Feature Engineering-3

연속형 특성

판매량

NULL\_index

Price\_index

Peak\_index

HOT\_index

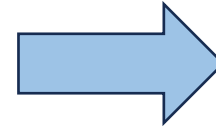
범주형 특성

대분류MDS

중분류MDS

소분류MDS

쇼핑몰MDS



Feature  
9개 선택



**Model**





# Model Architecture

9 features

	f1	f2	f3	f4	f5	f6	f7	f8	f9
1									
2									
3									

90  
days

...									
...									
...									
87									
88									
89									
90									

LSTM

$p$ : peak index  
 $\hat{y}$ : LSTM output

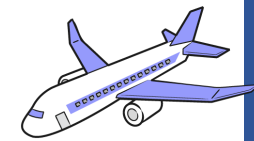
	$\hat{y}$	$p \cdot \hat{y}$
1		
2		
3		

MLP

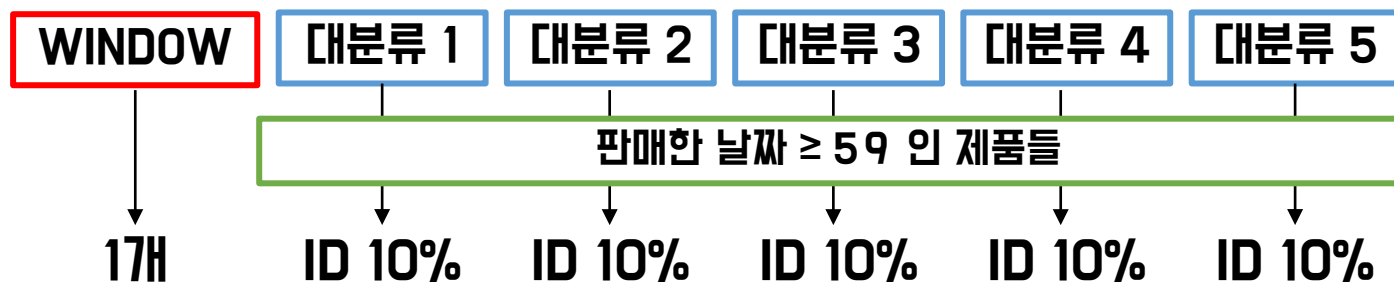
	$y^*$
1	
2	
3	

...		
18		
19		
20		
21		

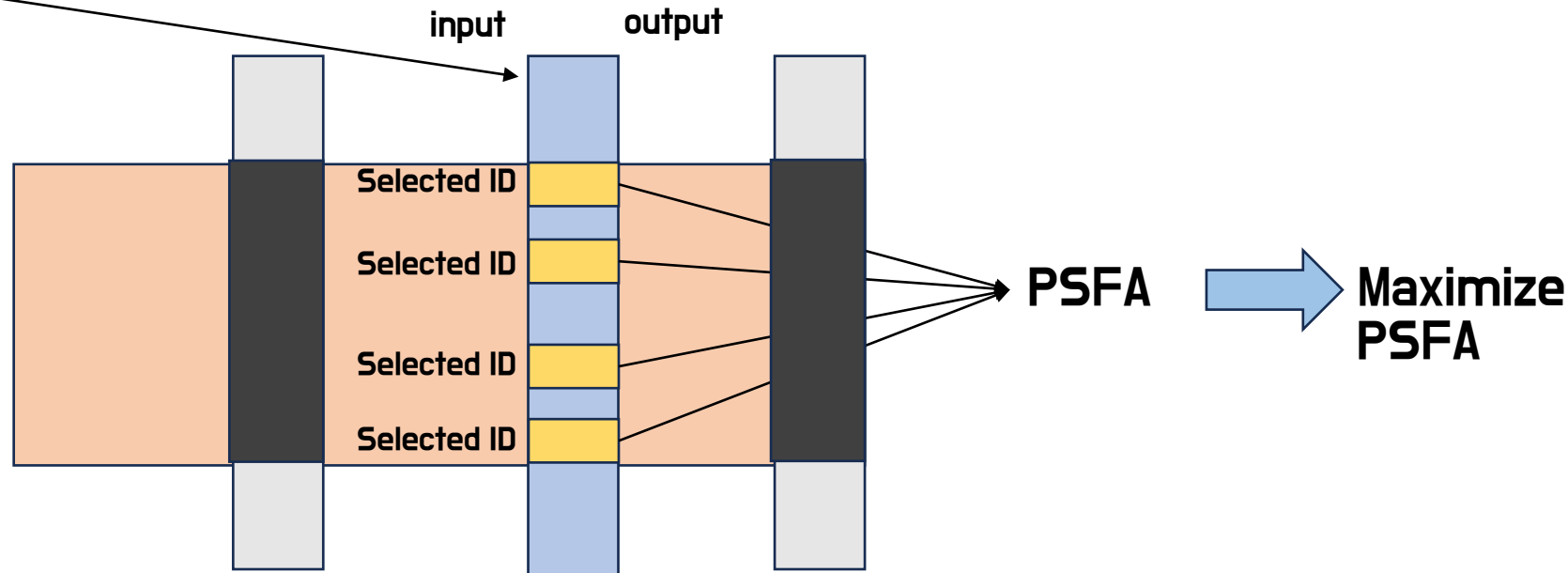
...		
18		
19		
20		
21		



# Train Method

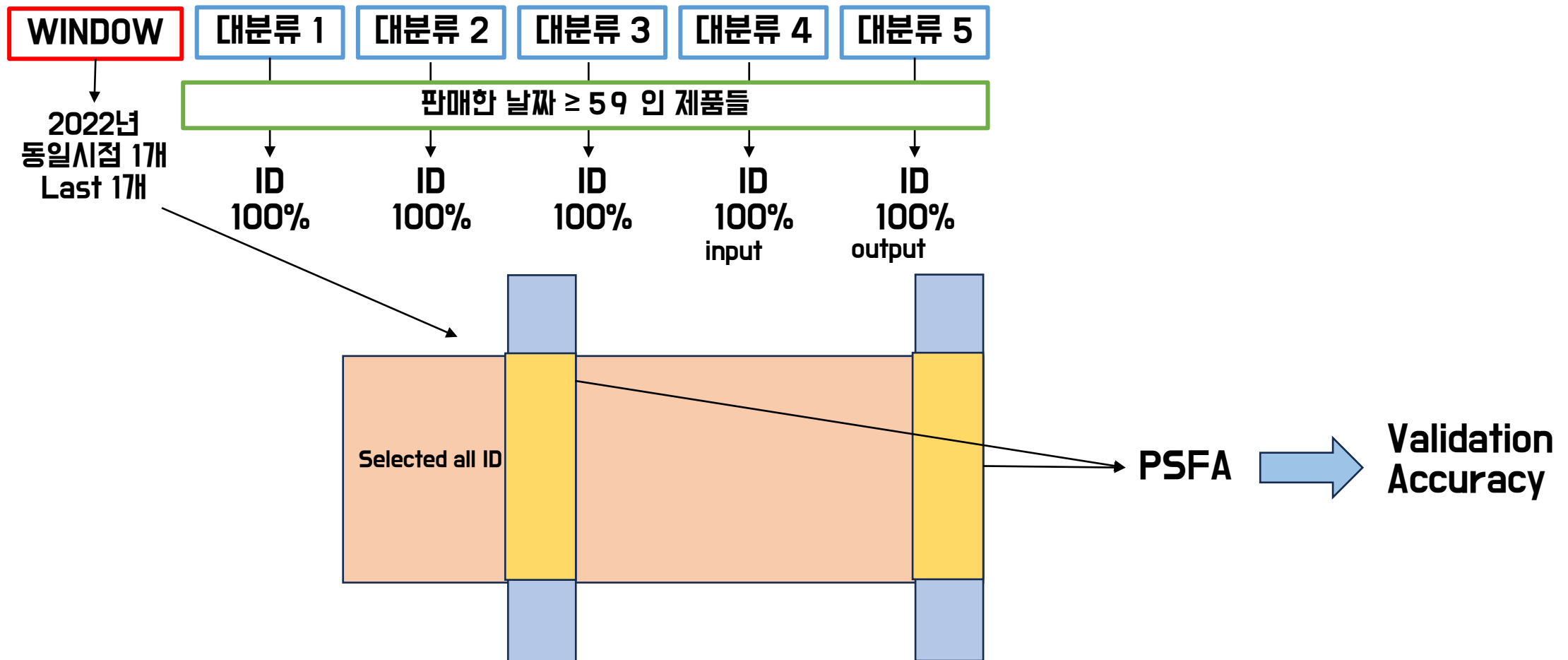


Mini-Batch  
Gradient Descent





# Validation Method



# Result

2

SME air



0.59749

47

2달 전

<b>Epoch</b>	<b>134</b>
<b>Train Accuracy (AVG)</b>	<b>0.61198</b>
<b>Validation Accuracy</b>	<b>0.60158</b>
<b>Private PSFA</b>	<b>0.59759</b>

LG AIMERS



SME AIR

