

하이테크 단기

[딥러닝응용실습] 머신러닝 기반 감성분석

한국폴리텍대학 성남캠퍼스 인공지능소프트웨어과
이혜정 교수

강의자 소개 : 이혜정

이 력

22.12 ~

- 한국폴리텍대학 성남캠퍼스 인공지능소프트웨어과

95.02 ~
22.01

삼성전자

- 인공지능 **임베디드/클라우드 SW** 개발 총괄
(자연어처리, 음성인식, 영상인식, 데이터분석)
- 인공지능/HCI **SW 및 HW** 선행 개발 총괄
(Interaction Device, Connectivity)
- 중장기 **기술 전략** 수립 (HCI, Robot, 스마트홈 등)
- C-Lab, 중소기업 컨설팅 (인공지능, 기술기획)
- 채용 관련 전문성/인성 면접 위원
- 한국/해외 등록 특허 발명 10건 이상

자격증

- 정보관리기술사
- 정보시스템 수석감리원

주요 업적

교육

- 소프트웨어 공학, 운영체제, 알고리즘
리눅스, 클라우드, AWS
- Java, Spring, Python
- 데이터분석, 인공지능

주요 과제 (삼성전자)

- TV용 **음성대화** 솔루션 (컨텐츠검색, 기기제어)
 - 7개 언어 음성인식 및 대화성공률 90% 이상
 - 100여개 다국어 음성인식 솔루션 탑재
- TV 연결기기 **자동인식** 및 통합제어
 - TV 연결 STB, 게임기, OTT 기기
 - 5여개국 250여개 사업자 STB 인식
- **TV 시청** 패턴 기반 선호채널 추출
 - 리모컨 동작 횟수 50% 이상 감소

강의 소개 : 머신러닝 기반 감성분석

■ 학습목표

- 감성 분석의 개념을 이해하고, 자연어처리와 머신러닝 기법 활용하여 감성 분석 구현
- 감성분석 Web App 개발 (문장 단위, 파일 단위)
- 감성분석 머신러닝 Tool 개발 (데이터 수집, 데이터셋 관리, 학습, 학습모델 관리)

■ 강의 및 실습 내용

- 텍스트 마이닝, 감성 분석의 이해
- [실습] 텍스트 전처리, 특징 추출
- [실습] 머신러닝/딥러닝 기반 감성분석
 - 특징 추출 (단어표현)
 - Supervised Machine Learning, Deep Learning, Transformer
- [MiniProject] Web App 및 Tool 개발
 - 데이터 수집 Tool
 - Flask 기반 감성 분석 Web App 개발
 - Flask 기반 감성 분석 머신러닝 Tool 개발

머신러닝 기반의 감성분석

텍스트 마이닝

텍스트 마이닝, Text Mining

- 텍스트에서 고품질 정보를 추출하는 과정
- 텍스트에 숨어 있는 패턴을 이용하여 어떤 사건을 예측하는 것
- 비정형의 텍스트 데이터로부터 패턴을 찾아내어 의미 있는 정보를 추출하는 분석 과정 또는 기법
- 자연어 처리 기법을 이용해 텍스트를 정형화된 데이터로 변환하고, 머신러닝 기법을 적용해 관심 있는 어떤 사건을 예측하고자 하는 방법론

텍스트 마이닝의 주요 적용 분야

■ 문서 분류, Text Classification

- 주어진 문서에 대해 미리 정의된 클래스로 분류하는 작업
- 스팸 메일 추출
- 감성 분석

■ 문서 생성, Text Generation

- 사람이 쓴 것과 유사한 문장을 만들어내는 작업

■ 문서 요약, Text Summarization

- 주어진 문서에서 중요하고 흥미있는 내용을 추출해 요약문을 생성하는 작업

■ 질의 응답, Question Answering

- 주어진 문장(context)을 읽고, 주어진 문제(question)에 대해 올바른 답(answer)을 생성하는 작업

■ 기계 번역, Machine Translation

- 한 언어로 작성된 문서를 다른 언어로 번역

■ 토픽 모델링, Topic Modeling

- 여러 문서에서 공통으로 등장하는 토픽(주제)을 추출하는 방법
- 다수의 문서에 잠재된 내용을 파악하는데 활용
- 각 토픽이 한 문서에서 어느 정도의 비중을 차지하는지와 같은 정보를 알 수 있음

머신러닝 기반 감성분석

감성분석의 이해

감성 분석, Sentiment Analysis

■ 감성(Sentiment) : 어떤 대상에 대한 주관적인 감정 (not facts)

- 보통 긍정(positive), 중립(neutral), 부정(negative)의 세 분류로 구분
 - 중립은 감정없이 객관적인 정보만을 전달하는 텍스트
 - 중립을 생략하고 긍정과 부정으로만 분류하기도 함
- 근래에는 긍정, 중립, 부정, 혼합으로 구분하기도 함

■ Sentiment = feelings

- Attitude
- Emotions
- Opinions

“The room was great, but the staff was unfriendly.”



```
graph LR; A[Opinion Mining] --> B

| Subject | Opinion    | Sentiment |
|---------|------------|-----------|
| Room    | Great      | Positive  |
| Staff   | Unfriendly | Negative  |


```

Subject	Opinion	Sentiment
Room	Great	Positive
Staff	Unfriendly	Negative

■ Opinion Mining

감성 분석, Sentiment Analysis

- 텍스트에 나타난 의견, 평가, 태도와 같은 주관적인 정보를 분석하는 것
 - 극성(Polarity) : 감성의 정도
 - 숫자로 표현 : (예) 0이면 중립, 양수이면 긍정, 음수이면 부정
 - Semantic Orientation 분석
 - Sentiment Content 추출, 식별
 - 감성 분석의 모호함, 어려움
 - 마블 영화는 액션이 화려하다
 - 마블 영화는 액션이 화려하여 좋다
 - 마블 영화는 액션은 화려하지만 내용이 부실하다
 - 영화가 안 좋다고 할 수 없다

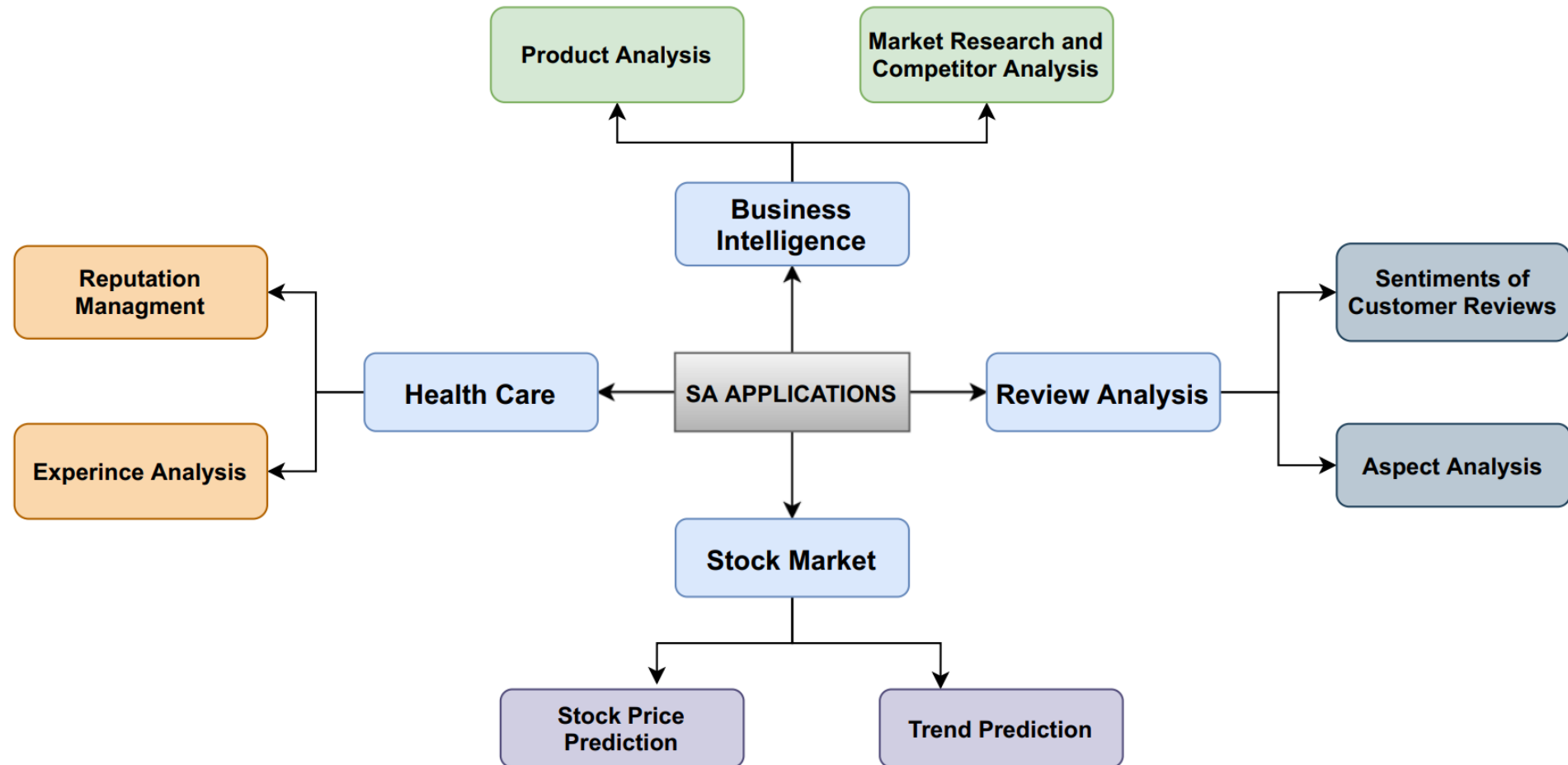
Sentiment Analysis 활용 질문

- 이 제품의 리뷰가 긍정적인가? 부정적인가?
- 고객에게 발송한 메일이 만족스러운가? 불만족스러운가?
- 광고 캠페인, 상품 출시에 대해서 고객들이 SNS에서 어떻게 반응하고 있는가?
- 당선된 대통령에 대한 블로거들의 attitude는 어떠한가?

Sentiment Analysis의 활용

- 제품 리뷰 중 부정적인 리뷰를 분석하여 개선 방안 도출
- 광고를 어디에 placement할 것인지 결정
- 고객과의 Q&A 중 Sentiment/Opinion이 있는 질의를 파악
- 정보 들 중에서 주관적인 정보를 추출하여 삭제
- 다양한 관점의 Summarization 생성
- 뉴스 기사의 편향적인 관점 식별

Sentiment Analysis의 활용



CSP가 제공하는 Sentiment Analysis

- NCP CLOVA Sentiment

- <https://www.ncloud.com/product/aiService/clovaSentiment>

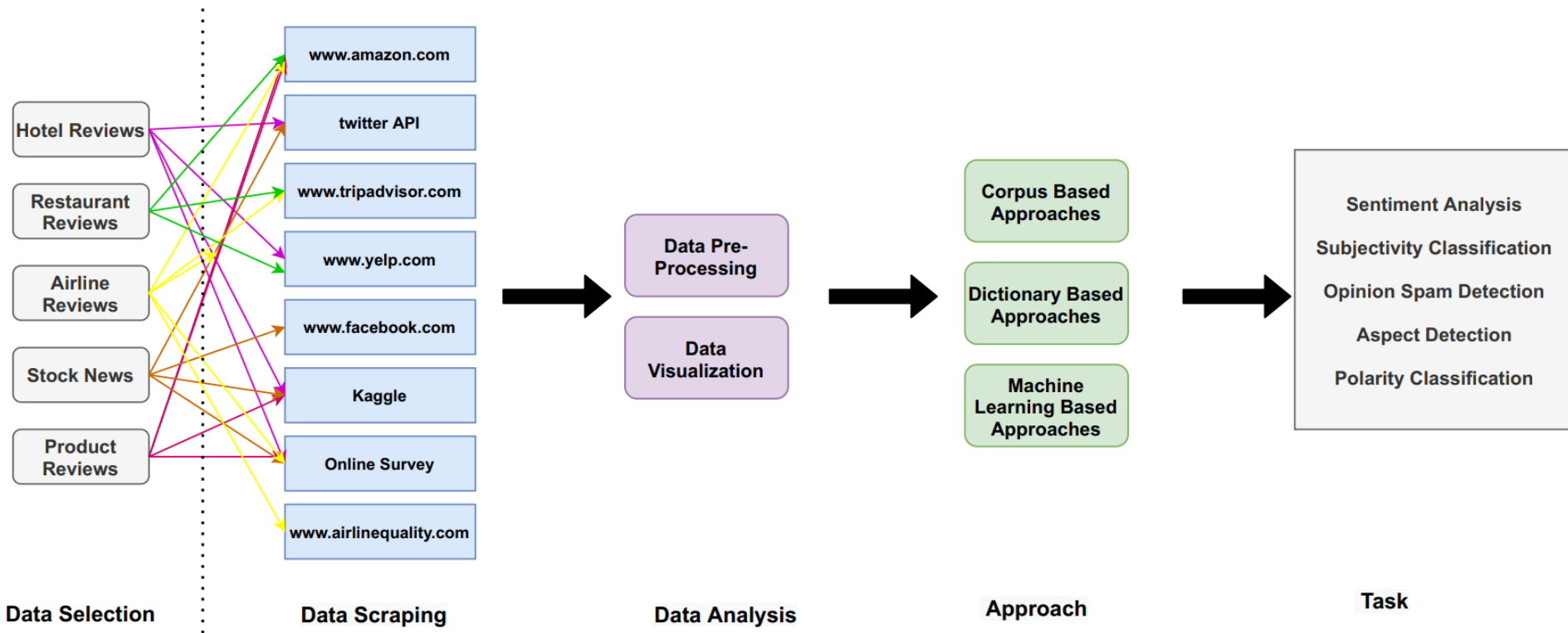
- Amazon Comprehend – Sentiment

- <https://docs.aws.amazon.com/comprehend/latest/dg/how-sentiment.html>

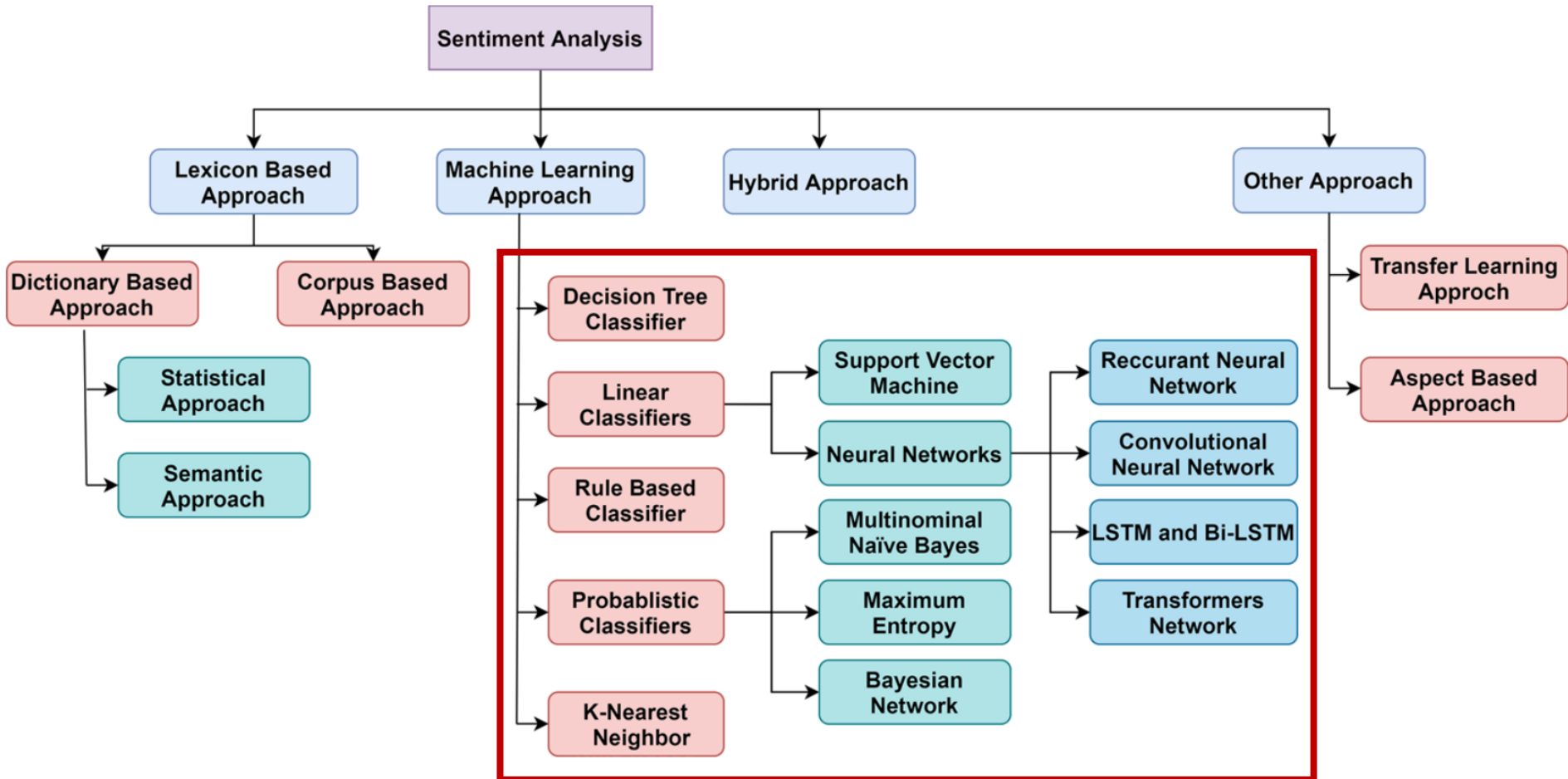
- MS Azure AI Services – Sentiment analysis wizard

- <https://learn.microsoft.com/en-us/azure/ai-services/language-service/sentiment-opinion-mining/how-to/call-api>

Sentiment Analysis 방법



Sentiment Analysis 방법



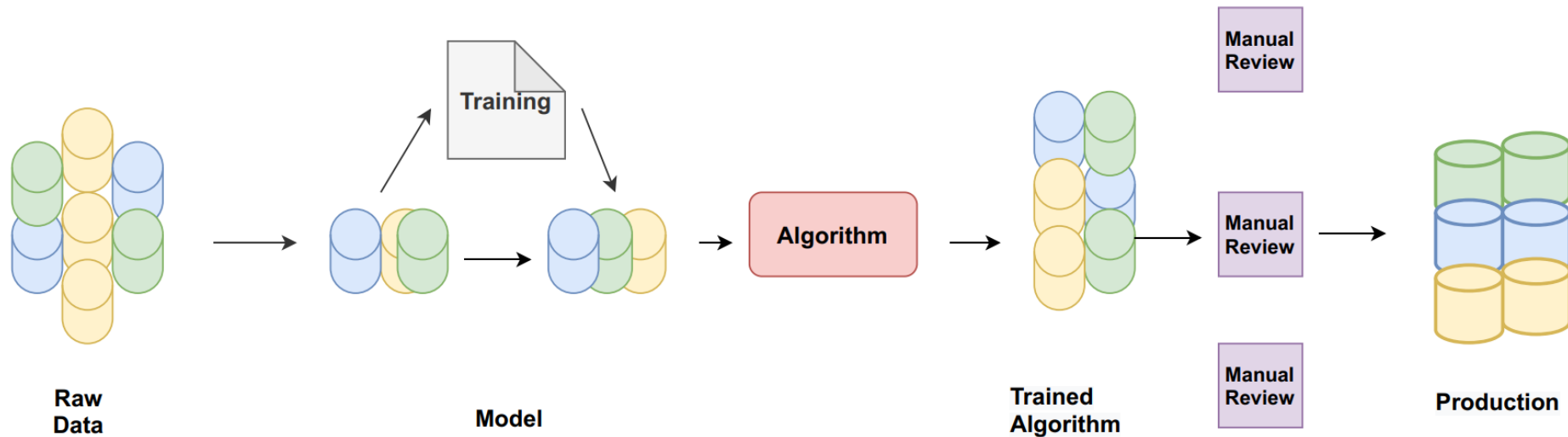
어휘(Lexicon) 기반의 감성 분석

■ 감성 사전에 기반하여 단어별 감성 점수를 계산하여 감성 분석

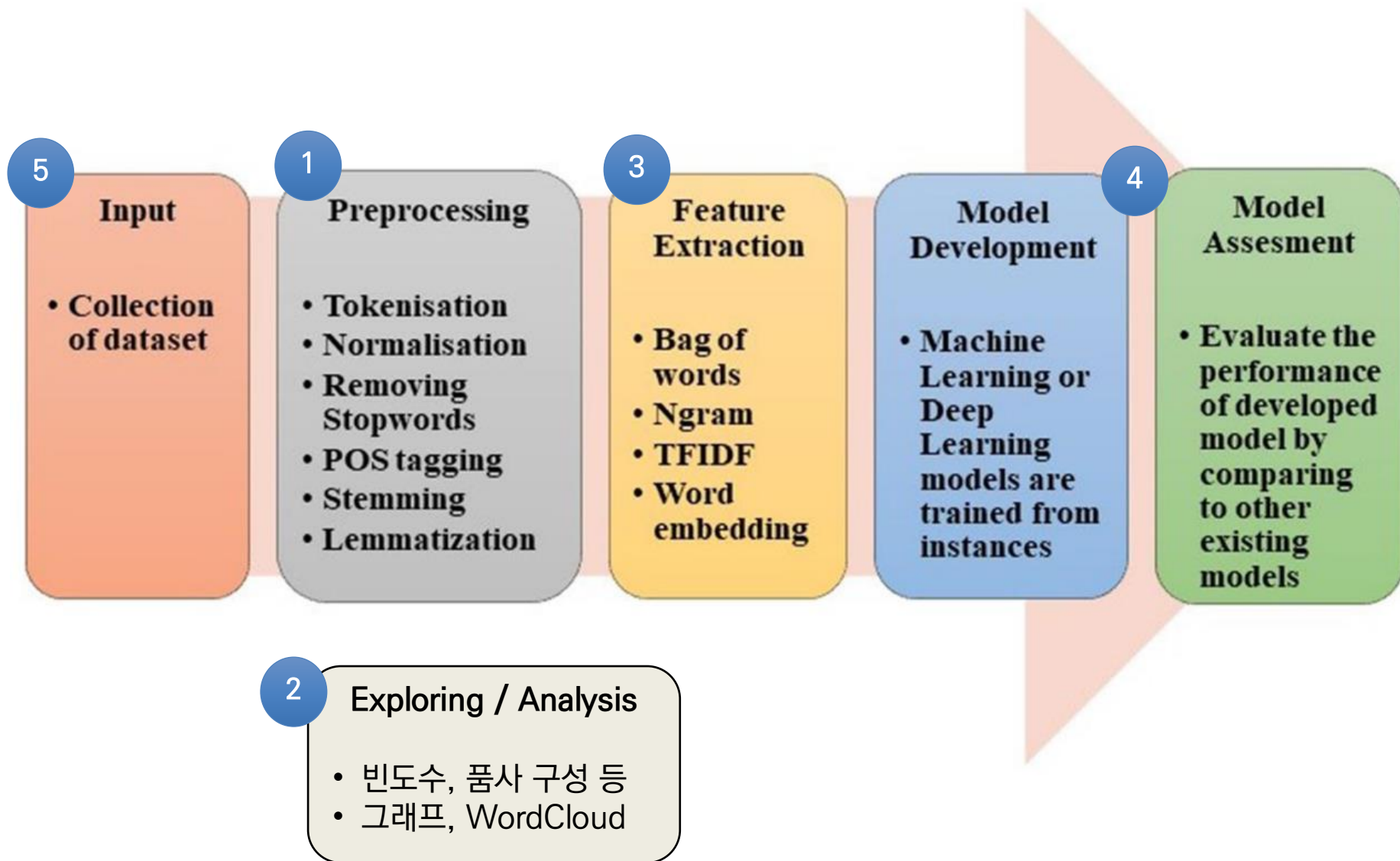
- 감성 사전 : 감성이 표현될 수 있는 명사, 형용사, 동사를 대상으로 모든 단어에 대해 긍정 혹은 부정의 감성을 붙여서 구축
 - Dictionary based Approach : 기존에 구축되어 있는 사전을 활용
 - 영어권 : SentiWordNet
 - 기존의 감성 사전 사용 시 적용하고자 하는 도메인에 적합하지 않을 수 있음
 - 예) 즐리다
 - Corpus based Approach : 도메인에서 문서를 모아서 Corpus(말뭉치)를 구축하고 이로부터 감성 사전을 구축하여 감성 분석
 - 많은 시간과 노력이 필요
- 마블 영화의 화려함과 빠른 전개가 시야를 어지럽게 한다.

Machine Learning 기반 감성 분석

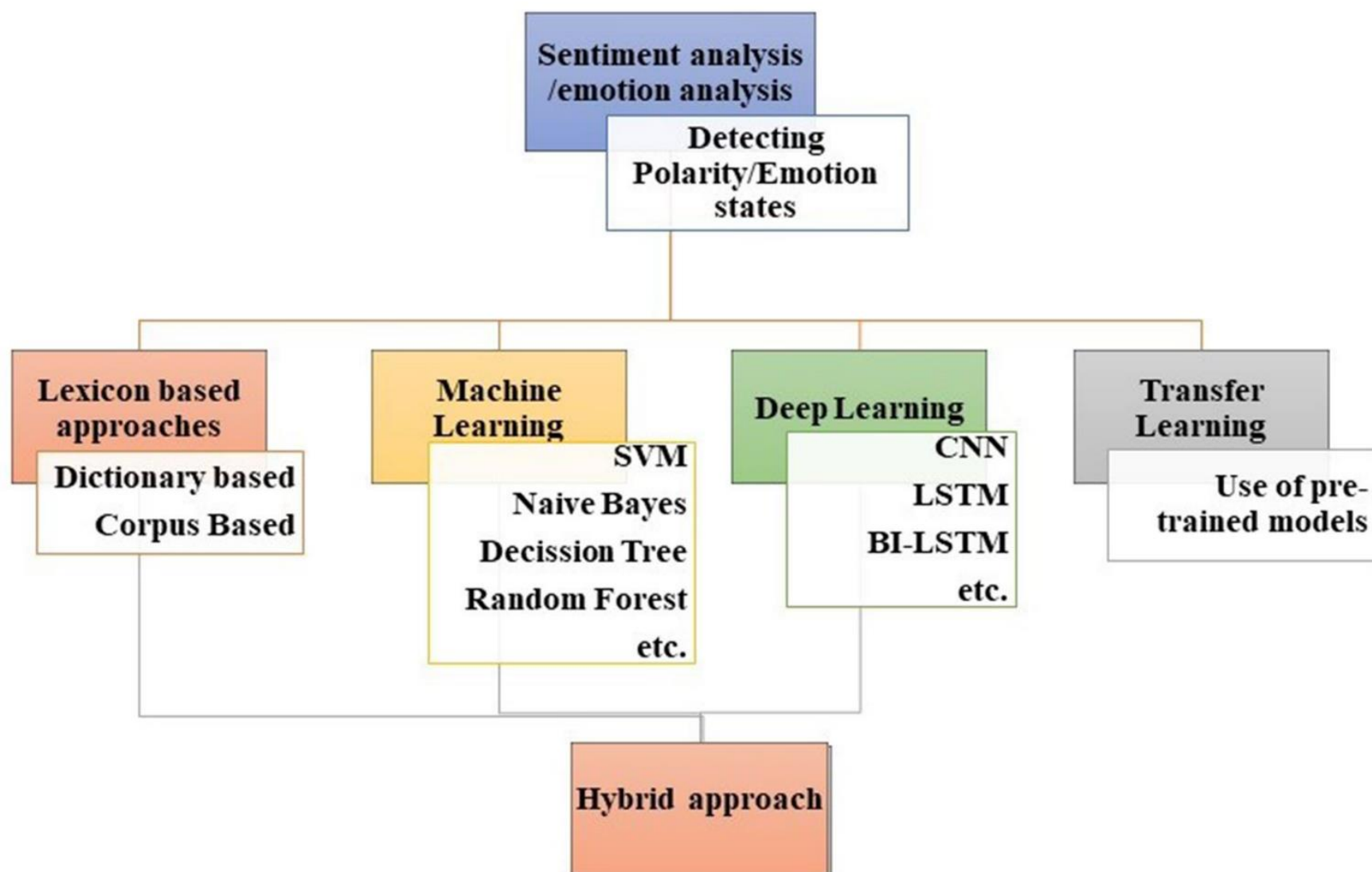
- 데이터셋으로부터 감성을 학습하여 새로운 데이터에 대한 감성 분석
- Supervised Machine Learning 기반 Sentiment Analysis
 - 데이터별 감성을 label하여 학습



감성 분석 절차



수고하셨습니다~~~~^



Supervised Machine Learning

