

중소기업 통계데이터 활용 정책아이디어 공모전

# TED를 활용한 기업 선별 모형



#Clustering

#Prediction

#TED



# CONTENTS

## 1 분석 배경

| 주제선정 배경, MBTI 성격 검사

## 2 분석 목적

| 제안아이디어

## 3 분석과정 및 결과

| Flow chart, Clustering, Prediction

## 4 결론

| 기존 정책과의 차이, 실현가능성, 기대효과

## 5 참고 자료 및 분석 도구

| 참고 문헌, 분석 도구

# 1 분석배경 주제선정 배경



## 중소기업이 필요로 하는 지원정책

좌측 이미지는 중소기업 관련 이슈 및 현황을 분석하기 위해 2017 ~ 2019년도의 '중소기업 정부지원' 관련 뉴스를 크롤링하여 워드클라우드를 생성한 결과이다.

이를 통해 중소기업이 지원받고자 하는 분야가 다양하다고 볼 수 있다. 따라서 **기업 특성에 따른 분류와 선별을 통한 맞춤형 지원**이 필요하다고 판단하였다.

## 현 중소기업 관련 지원 정책의 문제점

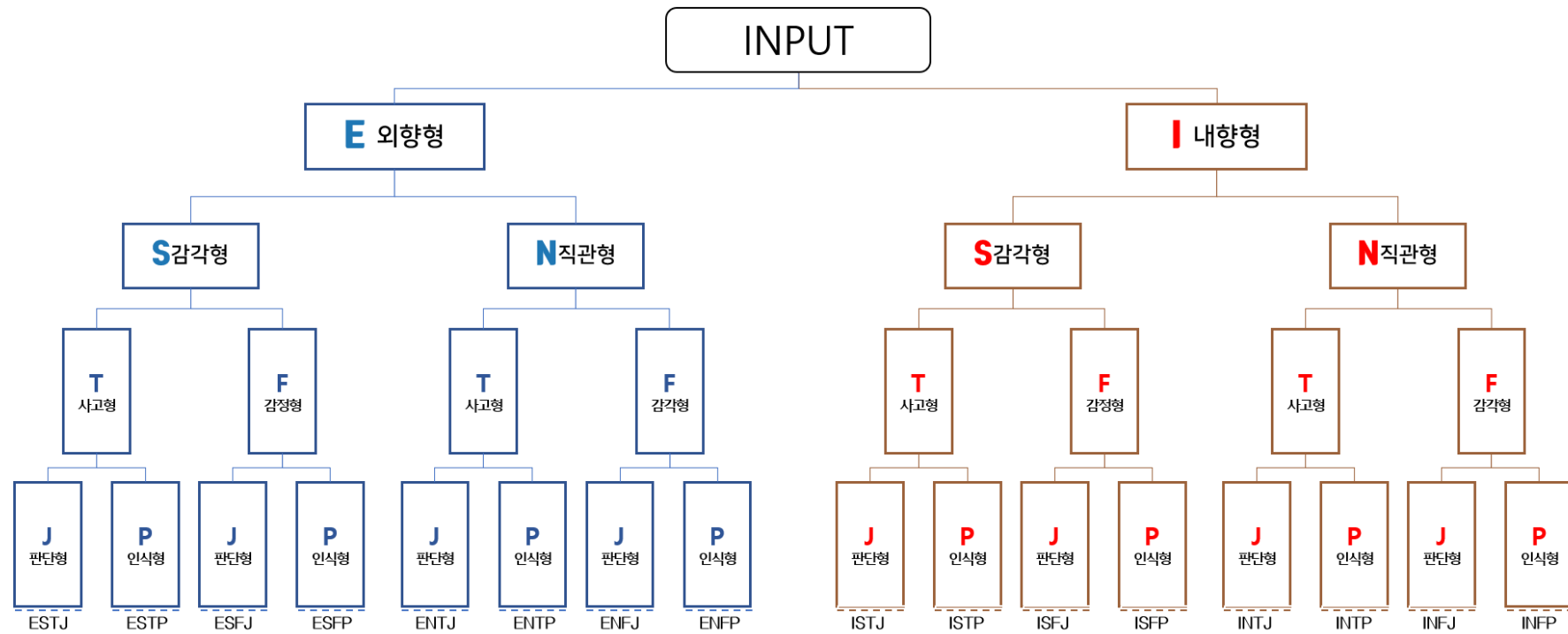
중소벤처기업부는 다양한 지원정책을 펼치고 있지만, 일각에서는 이러한 정책이 '자금지원'에만 초점이 맞춰져 있어, 기업의 산업구조를 반영하고 있지 못하다는 의견이 나오고 있다.

또한 '퍼주기식 지원'으로 좀비기업을 늘리는 정책이라는 비판 또한 생겨나고 있다.

이에 문제점을 해결하기 위한 데이터 분석 기반의 '**TED를 활용한 기업 선별 모형**'을 제안한다.

# 1 분석배경

## MBTI 성격 검사



4가지 분류 기준에 따른 결과에 의해, 대상자를  
16가지 성격 유형 중의 하나로 분류하는 검사

## 2 분석 목적 제안 아이디어

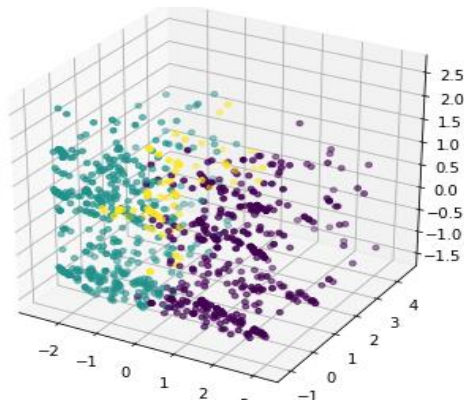


TED는 3단계의 분류 과정을 거쳐 기업을 특성별로 분류한다.

1. 산업군 (T) 2. 기업성장단계 (E) 3.기업의 특징 (D)

분류과정을 거친 기업을 예측분석을 실시하여 우수기업을 선정한다.

# 3 분석 과정 및 결과 FLOW CHART



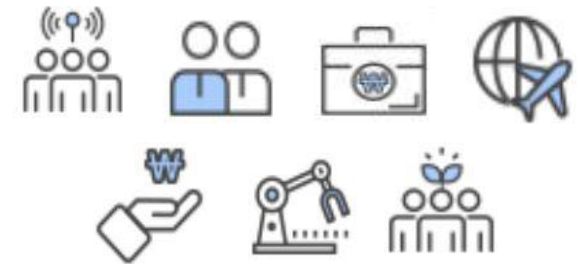
## 군집분석

Clustering 알고리즘을 통한  
기업 특성 군집 생성



## 예측분석

머신러닝 알고리즘을  
이용한 성장예측 기업 선별



## 기업선별 및 지원

두 가지 분석을 통해 선정된 기업의  
특성을 고려한 지원

# 3 분석 과정 및 결과 활용 데이터

활용변수	기존 문항	변경 문항
기술개발동기 1순위 (A1S1)	1. 내수시장 점유율 확대 유지 ⋮ 7. 기타	1.시장확대형 (1,2) 2.다양성확대형(3) 3.개선형 (4,5,6) 4.기타
기술개발목적 1순위 (A1N1)	1.시장점유율 확대 ⋮ 15.기타	1.시장확대형 (1,2) 2.다양성확대형(10,11) 3.개선형(4,5,6,8,9,12) 4.인력형 (3,7) 5.기타 (13,14,15)
향후 1년간 중점 투자계획 분야 (C2S2)	1.신제품 개발 ⋮ 4.기존공정 개선	1.신제품 개발 (1) 2.기존 개선 (2,4) 3.신 공정 개발 (3)

## 기술 통계조사

탐색적 자료 분석(EDA) 실시  
12개 변수 선택 및 각 변수 문항 축소

활용변수
매출 (INC_A_41)
매출원가 (INC_A_42)
급여 (INC_A_45)
복리후생비 (INC_A_47)
임차료 (INC_A_48)

2017년 독립변수



영업손익 증가여부	$\frac{2018\text{년 영업손익} - 2017\text{년 영업손익}}{2017\text{년 영업손익}}$
-----------	---

예측변수  

$$(INC\_A\_21 - INC\_A\_61) / (INC\_A\_61)$$
  
 (0: 증가하지 못함, 1: 증가함)

## 실태조사

손익계산서의 19개 항목 독립변수 선택  
영업손익 증가여부 종속변수 선정



# 3 분석 과정 및 결과 군집분석

처리방법	Spectral	DBSCAN	AGG
Original	0.636	0.550	0.601
Dummy	0.659	0.480	0.567
PCA	0.588	0.609	0.627
Dummy +PCA	0.637	0.573	0.542

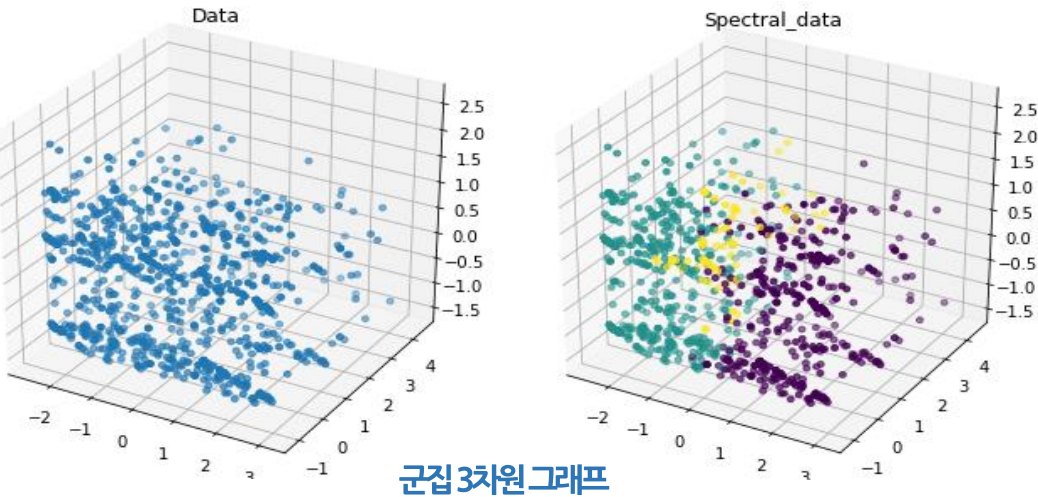
Silhouette coefficient

## Spectral Algorithm

유사성 매트릭스의 스펙트럼(고유값)을 사용한 데이터 차원축소 후, 군집화 하는 알고리즘이다.

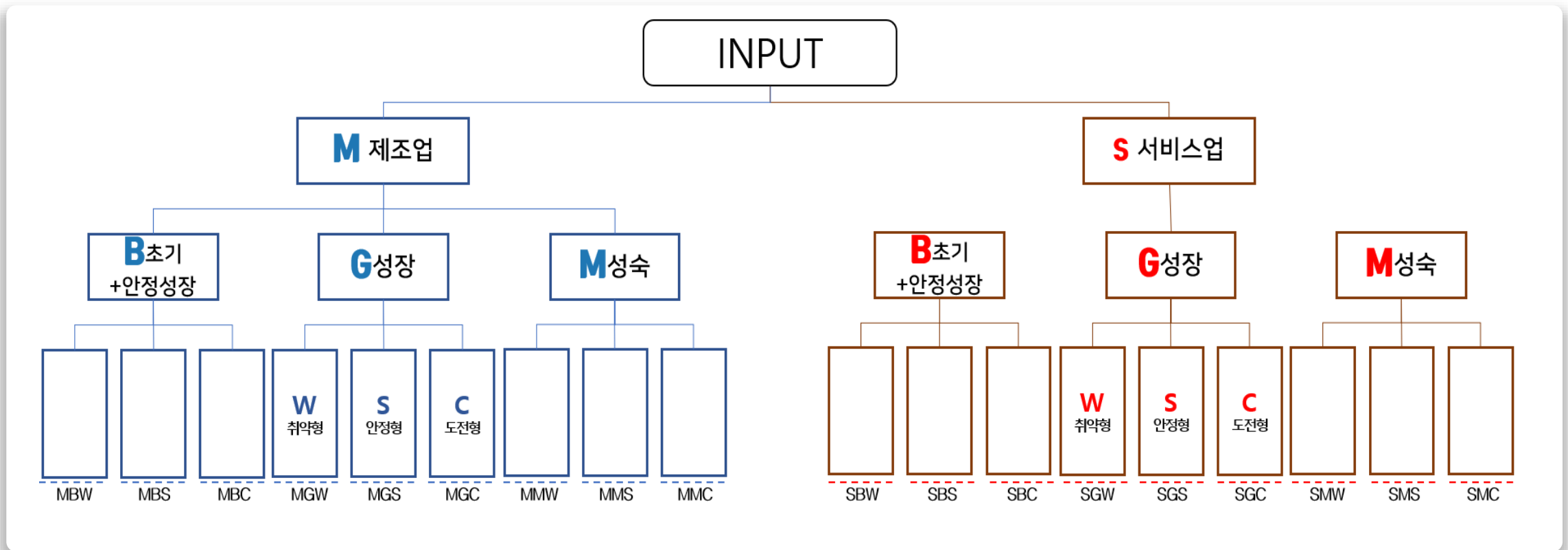
전처리를 완료한 기술통계 자료를 효율적으로 활용하기 위해, Dummy, PCA 등의 변환 과정을 진행하였다.

그 후, Silhouette coefficient 비교를 통해 최종적으로 Dummy 변수 처리한 Spectral Algorithm을 선택하였다.





# 3 분석 과정 및 결과 군집분석



## 분석결과

그룹1 (취약형, Weak): 자금지원과 기술개발에 대한 지원 필요

그룹2 (안정형, Stable): 시장확대와 개선에 대한 지원 필요

그룹3 (도전형, Challenge) : 다양성 확대를 위한 자금지원 필요

# 3 분석 과정 및 결과

## 예측분석

알고리즘	Accuracy(정확도)
XGBoost Classifier	0.7318944
GBM Classifier	0.6211031
Logistic Regression	0.5827998

Accuracy Score

TED 그룹	성장 기업 비율
MMC (제조, 성숙, 도전)	43% (87/201)
MMW (제조, 성숙, 취약)	40% (77/191)
MGW (제조, 성장, 취약)	37% (49/132)

실제 TED 모형 결과

## XGBoost Algorithm

오차를 학습하는 Boosting 계열 알고리즘 중, 트리 기반의 알고리즘 학습에서 우수한 성능을 보이는 알고리즘이다.

종속변수 ‘영업손익 증가여부’ 예측을 목표로 3개의 알고리즘 후보군을 두고, Accuracy Score를 통해 평가를 진행하였다.

73%의 가장 높은 정확도를 보여준 XGBoost Algorithm을 선택하였다,

## 실제 데이터 예측 결과

TED 모형을 이용해 나누어진 18개의 군집 중 기업의 수가 가장 많은 상위 3개 군집의 결과이다.

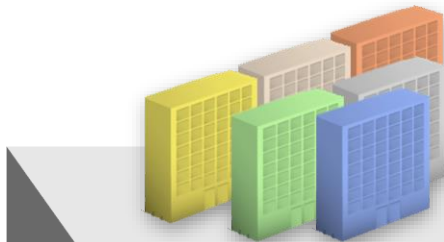
각 군집 별 성장이 예측되는 기업 비율은 평균 40%대이다.  
 (성장기업 비율 = 군집 별 성장 기업 / 군집 별 전체기업)  
 이를 통해, 성장 예측 기업을 선별할 수 있다.

# 3 분석 과정 및 결과 시나리오

STEP1  
중소벤처기업부  
지원사업에 신청하기



STEP2  
기업 분류를 위한 TED  
기술 통계 조사 참여하기



STEP3  
TED 모형을 통해  
배정된 군집 확인하기



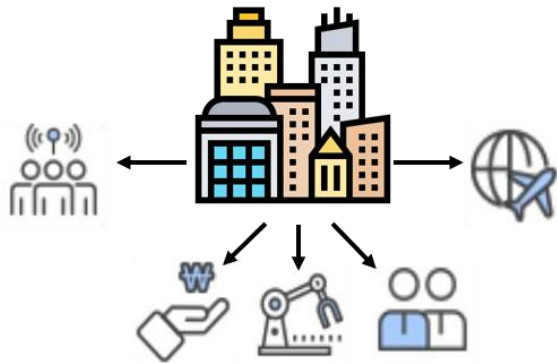
STEP4  
기업 성장예측 모델을 통해  
지원 가능 여부 확인하기



STEP5  
군집 특성을 고려한  
맞춤 지원 받기



# 4 결론 기존 정책과의 차이



## 맞춤지원

각 기업의 특징을 고려해 맞춤 지원 정책을 시행할 수 있다.



## 지속가능성

기업의 지속가능성 평가를 통한 지원으로 미래 성장가능성이 높은 기업을 발굴해내는데 도움을 줄 수 있다.



## 정확한 기준 정의

데이터 분석을 활용하였기에 정확한 기준 정의가 가능하다.

# 4 결론 실현 가능성



## 데이터를 이용한 기업선정

해당 아이디어의 핵심 알고리즘인 군집분석(Spectral model)의 실루엣 계수는 0.659, 예측분석(XGboost)의 정확도는 0.73이다.

이는 'TED'모형을 신뢰할 충분한 수치이다. 또한 모든 과정은 공모전 내 제공된 데이터를 통해 진행된 것이므로,

같은 형식의 더욱 많은 데이터가 수집된다면 아이디어의 실현가능성이 더욱 높아질 것이다.

## 제주시 빅데이터 지원 사업

최근 제주시의 '중소기업 빅데이터 지원 사업'이 일자리 창출과 매출 증대에 효과를 보이고 있다. 해당 사주업은 도 내 중소기업들의 기업 내부 데이터 및 사회 데이터 분석을 통해 기업혁신과 새로운 비즈니스 창출을 지원한다.

이처럼 빅데이터 분석 기반의 기업 지원 사업은 현실적으로 도움을 주고 있으며, 본 문서의 TED모델 실현 가능성을 기대한다.

# 4 결론 기대효과

## 자원 낭비 절감

데이터분석을바탕으로기업에게  
실질적으로필요한정책을파악할수있다.  
이를통해낭비되는자원을절약할수있다.



## 적시성 효과

데이터를 활용한 기업의 실시간  
평가가 가능해지므로 적시성  
효과를 기대할 수 있다.

## 효율성

기존의 까다로웠던 기업 심사 절차를  
축소시켜, 빠르고 정확한  
기업선정에 도움을 줄 수 있다.



# 5 결론 분석도구 및 참고자료

## 분석 도구



## 참고 자료

- \* 김선우, 김재원 (2020). 혁신성장을 위한 중소기업 R&D지원 개선방안
- \* 이양현, 심상규 (2007). 중소기업의 성장단계 판별모형에 관한 연구
- \* 최영수, 장욱재무(2007). 변수와 주가를 결합한 상호저축은행의 부실예측모형
- \* Stephen Roberts etc 3 (2020). Machine Learning-Based Financial Statement Analysis

