

“웹 기반 고차원 데이터 시각화 툴”

2018-1 Media Project | Final Presentation

Interactive Visualization Tool:

MDS-Classifier

“송용승 이동영 정예지”

CONTENTS

01. 프로젝트 개요

02. 프로젝트 구현

03. 참고 논문

04. 차후 계획

Contents 01.

프로젝트 개요

팀원 및 역할

- **송용승** – 소프트웨어학과
역할: UX, UI 및 웹 디자인 기획
- **이동영** – 미디어학과
역할: 기술 연구 및 SW 개발
- **정예지** – 미디어학과
역할: 기술 연구 및 SW 개발



Contents 01.

프로젝트 개요

프로젝트 기획 배경

01. Industry 4.0

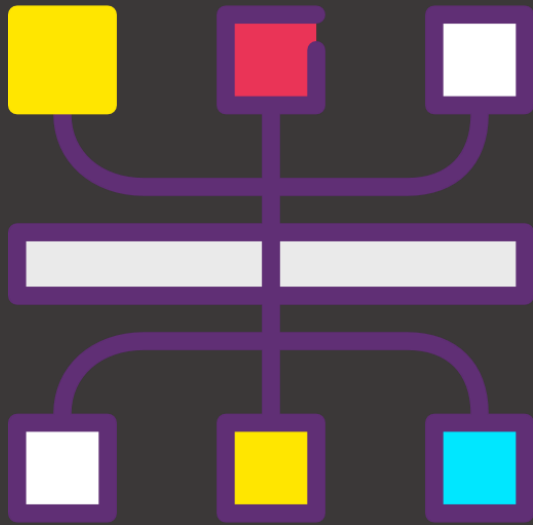
- 학문 및 기술의 경계가 없어지고 여러 분야의 기술이 융합되어 새로운 기술 혁신
- 획기적인 기술 진보, 파괴적 기술에 의한 산업재편, 전반적인 시스템의 변화



Contents 01.

프로젝트 개요

프로젝트 기획 배경



02. Machine Learning

- 인공지능, 빅데이터, 사물인터넷 등 기술이 부상
디지털 기기와 인간의 융합으로 파급력 급속도 진화
- 특히 인공지능의 하위 영역인 머신 러닝은 이미지 인식
자연어 처리 등 인공지능 영역의 많은 부분에 기여

Contents 01.

프로젝트 개요

프로젝트 기획 배경

03. Multi Dimensional Data

- 머신 러닝 모델의 성능에 영향을 미치는 차원 문제 발생
- 고차원 데이터 셋에서 차원 축소, 확장 등 문제 해결을 위한 많은 연구 진행



Contents 01.

프로젝트 개요

프로젝트 기획 배경



04. Data Visualization

- 차원 문제를 해결하기 위해 데이터 시각화 기법을 사용
- 머신 러닝으로 부족한 데이터에 대한 이해를 돕고, 고차원 데이터에서 발생하는 차원 문제에 대한 기여

Contents 01.

프로젝트 개요

프로젝트 기획 배경

01. Industry 4.0

- 학문 및 기술의 경계가 없어지고 여러 분야의 기술이 융합되어 새로운 기술 혁신
- 획기적인 기술 진보, 파괴적 기술에 의한 산업재편, 전반적인 시스템의 변화

02. Machine Learning

- 인공지능, 빅데이터, 사물인터넷 등 기술이 부상
디지털 기기와 인간의 융합으로 파급력 급속도 진화
- 특히 인공지능의 하위 영역인 머신 러닝은 이미지 인식
자연어 처리 등 인공지능 영역의 많은 부분에 기여

03. Multi Dimensional Data

- 머신 러닝 모델의 성능에 영향을 미치는 차원 문제 발생
- 고차원 데이터 셋에서 차원 축소, 확장 등
문제 해결을 위한 많은 연구 진행

04. Data Visualization

- 차원 문제를 해결하기 위해 데이터 시각화 기법을 사용
- 머신 러닝으로 부족한 데이터에 대한 이해를 돕고,
고차원 데이터에서 발생하는 차원 문제에 대한 기여

Contents 01.

프로젝트 개요

프로젝트 기획 배경

“4차 산업혁명”으로 인해 새로운 기술 혁신이 일어났고,
인공지능, IoT, 빅데이터 등의 기술이 비약적인 발전을 이루었다.

특히, 머신 러닝은 이미지 인식, 자연어 처리 등 많은 분야에서 기여하였고 현재도 발전 중이다.

이러한 머신 러닝의 모델 성능에 가장 큰 영향을 미치는 것이 데이터의 차원 문제이다.

데이터의 차원이 높아질 수록 차원의 저주 문제가 생겨 모델의 성능이 급격하게 저하된다.

이러한 고차원 데이터의 차원 축소, 증가 문제를 해결하기 위한 많은 방법이 있다.

우리는 이러한 고차원 데이터의 차원 문제를 데이터 시각화 기술 측면에서 접근하여 해결하고자 한다.”

“본 프로젝트는 보다 나은 시각화 툴을 개발하기 위해
두가지 사안에 방점을 두었다.”

Contents 01.

프로젝트 개요

기획 강조점

1. 고차원 데이터 셋에서의 데이터 분류 및 탐색으로 의사결정 문제 해결

- 데이터의 패턴을 한눈에 파악하고 탐색할 수 있을 것으로 기대된다.
- Machine Learning 실무자와 데이터 분석가 모두에게 유용한 도구를 제공하고자 한다.
- 사용자와의 상호작용을 통해 의사 결정 문제에 도움을 주는 데에 기여하고자 한다.

2. 직관적인 Data Visualization 인터페이스

- 직관적인 인터페이스를 통해 쉽고 간단하게 사용자와 상호작용하는 시각화 앱
- 섹션 별 플로우를 통한 직관적인 사용법과 즉각적인 분석 결과 도출

Contents 01.

프로젝트 개요

기획

웹 기반 고차원 데이터 시각화 툴

MDS-Classifler

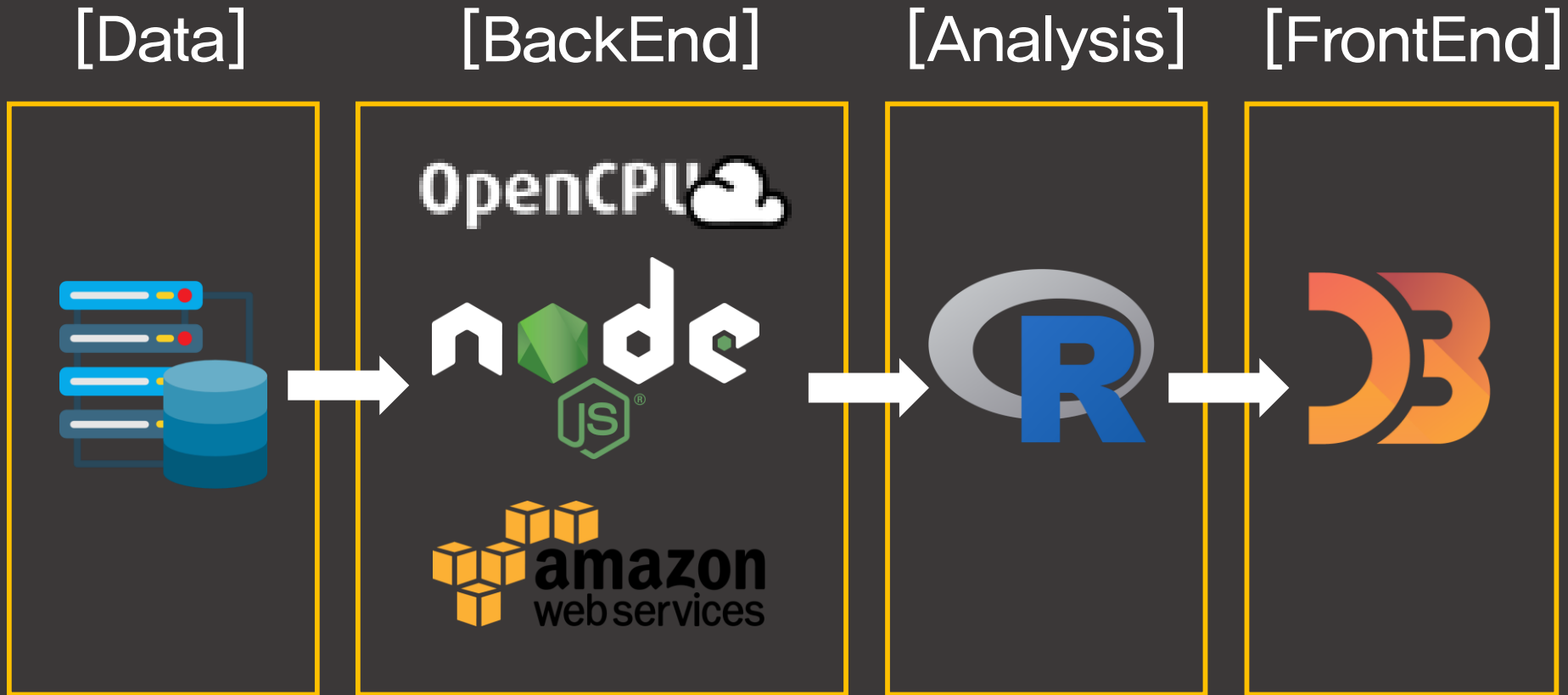


- 인터랙티브 웹 기반 시각화 도구
- 자신이 원하는 데이터 셋 분석 가능
- 고차원 데이터의 분류 및 탐색을 위한 다양한 기능 제공
- 차원 그룹간 상관성 분석 기능 (CCA)
- 데이터에 대한 이해, 실무자 및 분석가에게 통찰력 제공

Contents 02.

프로젝트 구현

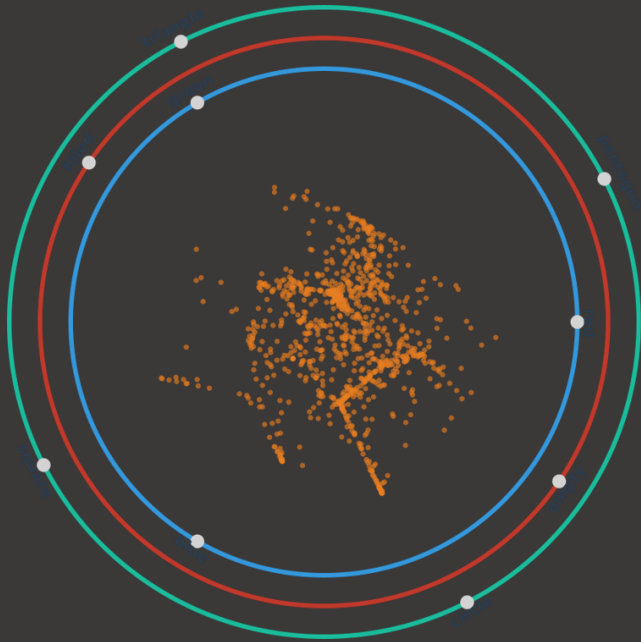
구현 프로세스



Contents 02.

프로젝트 구현

적용 기술



01. 동심원 시각화 – Concentric RadViz

- 다 차원의 원에 속성 값을 할당하여 데이터를 탐색할 수 있다. 최대 6개의 차원까지 생성하여 시각화 할 수 있다.
- Ono, J. H. P., Sikansi, F., Corrêa, D. C., Paulovich, F. V., Paiva, A., & Nonato, L. G. (2015, August). Concentric RadViz: visual exploration of multi-task classification. In Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on (pp. 165–172). IEEE.

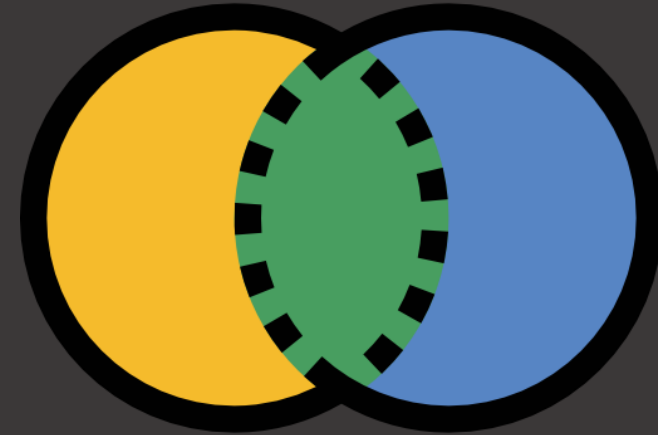
Contents 02.

프로젝트 구현

적용 기술

02. 변수 결합 – Feature Combination

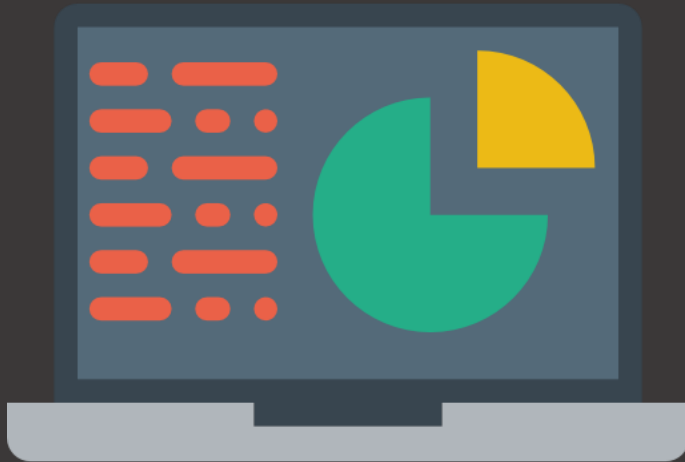
- 원으로 표현된 RadViz의 차원에 할당된 Feature의 위치를 변경함으로써 변수의 결합을 시각화 측면으로 해결하였다.
- Exploring the Best Classification from Average Feature Combination – Jian Hou,¹ Wei-Xue Liu,¹ and Hamid Reza Karimi²



Contents 02.

프로젝트 구현

적용 기술



03. 알고리즘 및 통계 기법 적용

- RadViz 시각화 알고리즘과 Feature를 할당 시키기 위한 알고리즘을 적용하였다.
- 각 다차원 동심원 그룹에 할당된 Feature를 집단 변수로 묶어 그룹 간 상관성 분석을 실시하였다. (CCA)

Contents 02.

프로젝트 구현

적용 기술

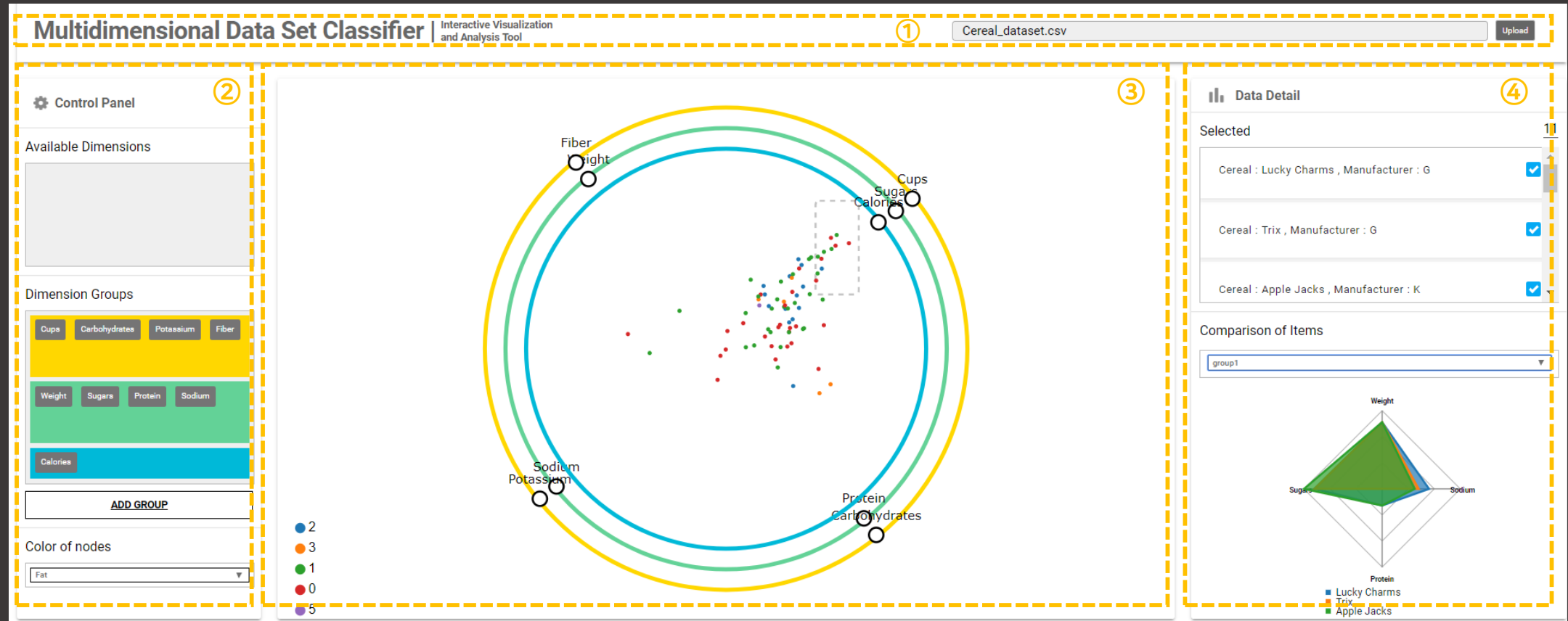
04. 기타 시각화 기능 제공

- 범주형 변수를 통한 노드 색 구분
- 노드의 개별 데이터를 레이더 차트를 통한 비교
- Sigmoid활성 함수를 통한 차원 내 Feature에 선택적 가중치 부여



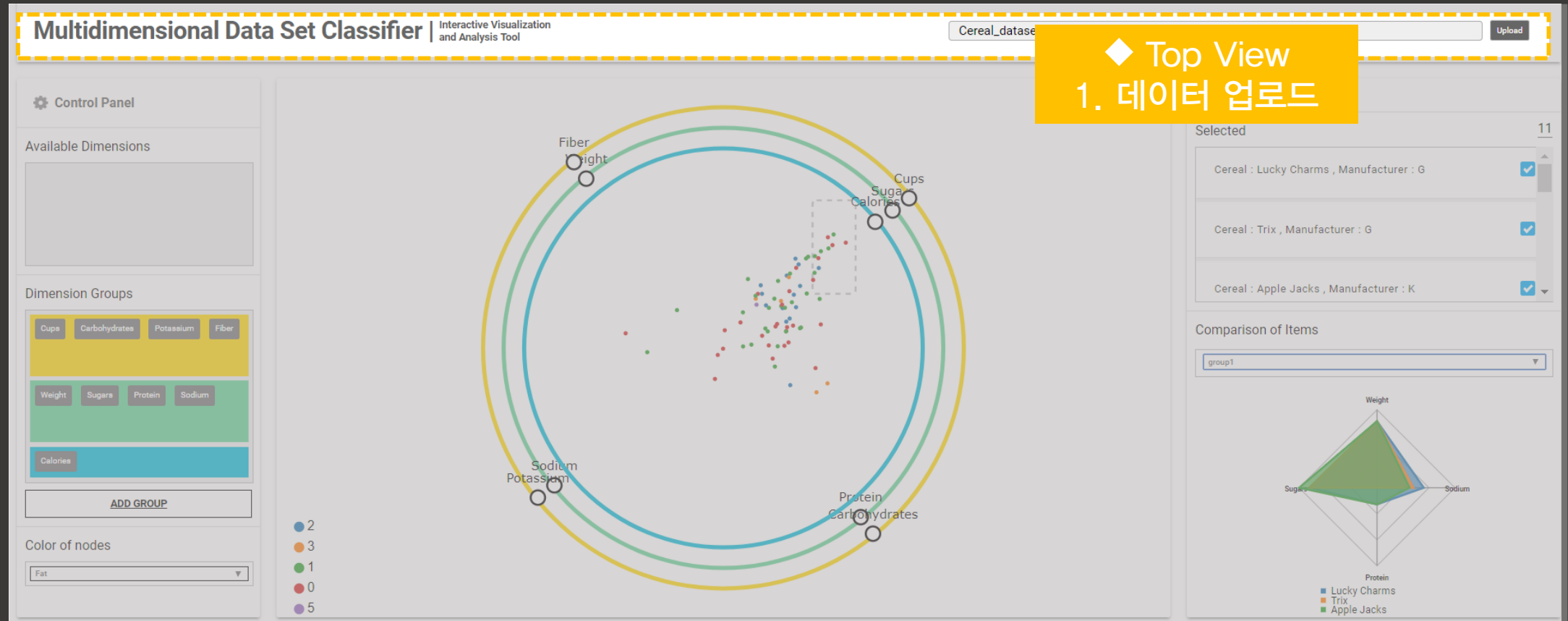
Contents 02.

프로젝트 구현 기능 설명



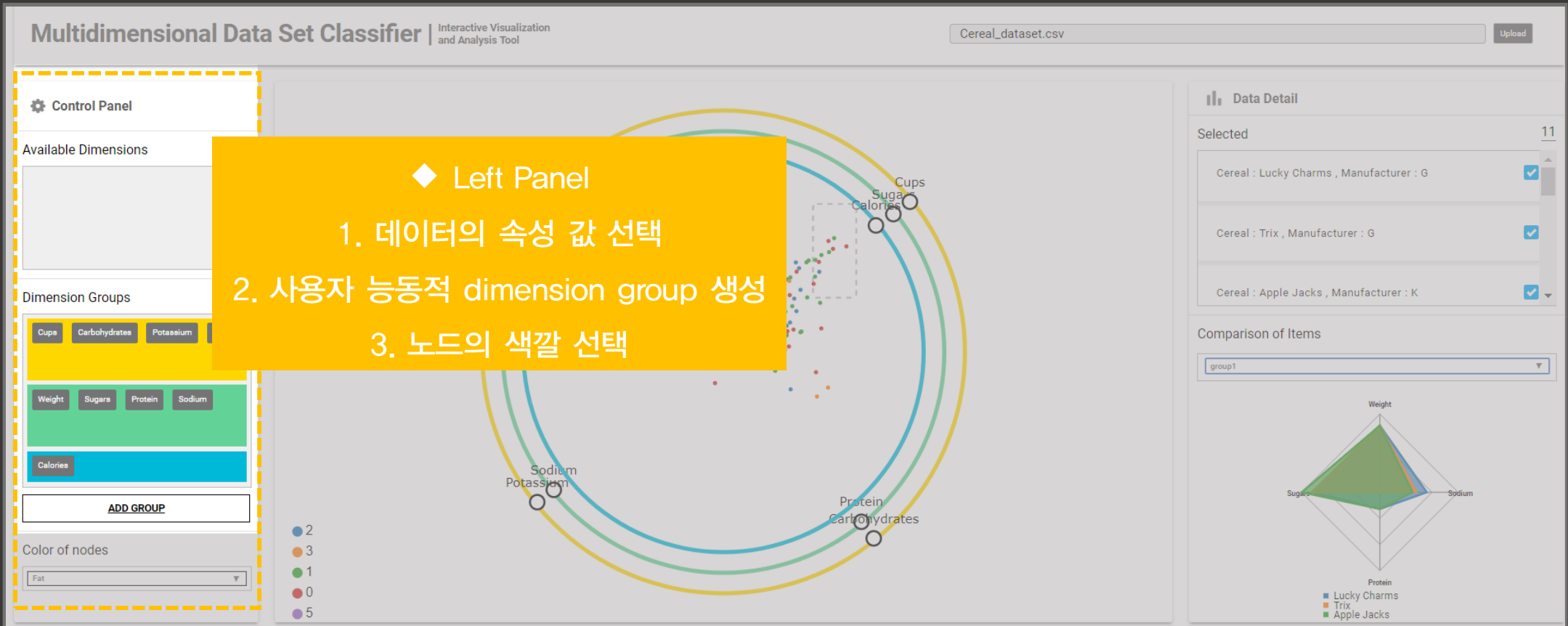
Contents 02.

프로젝트 구현 기능 설명



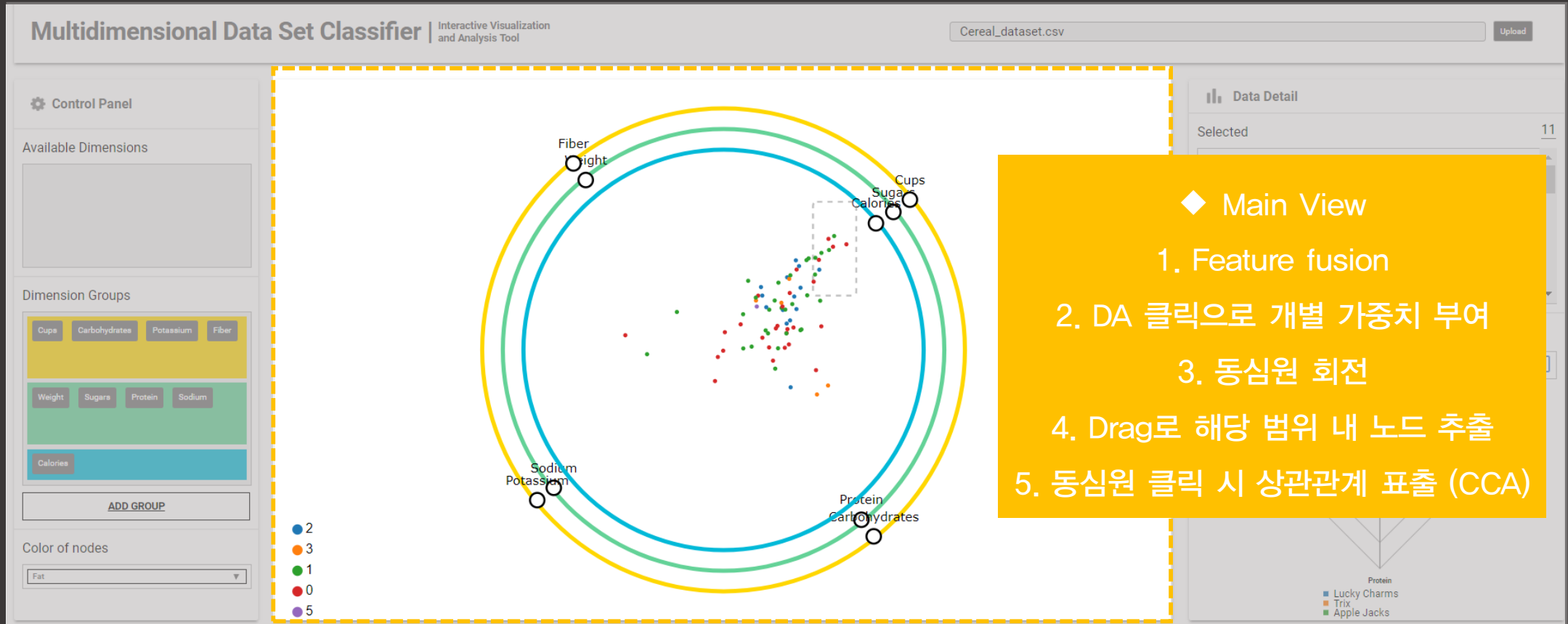
Contents 02.

프로젝트 구현 기능 설명



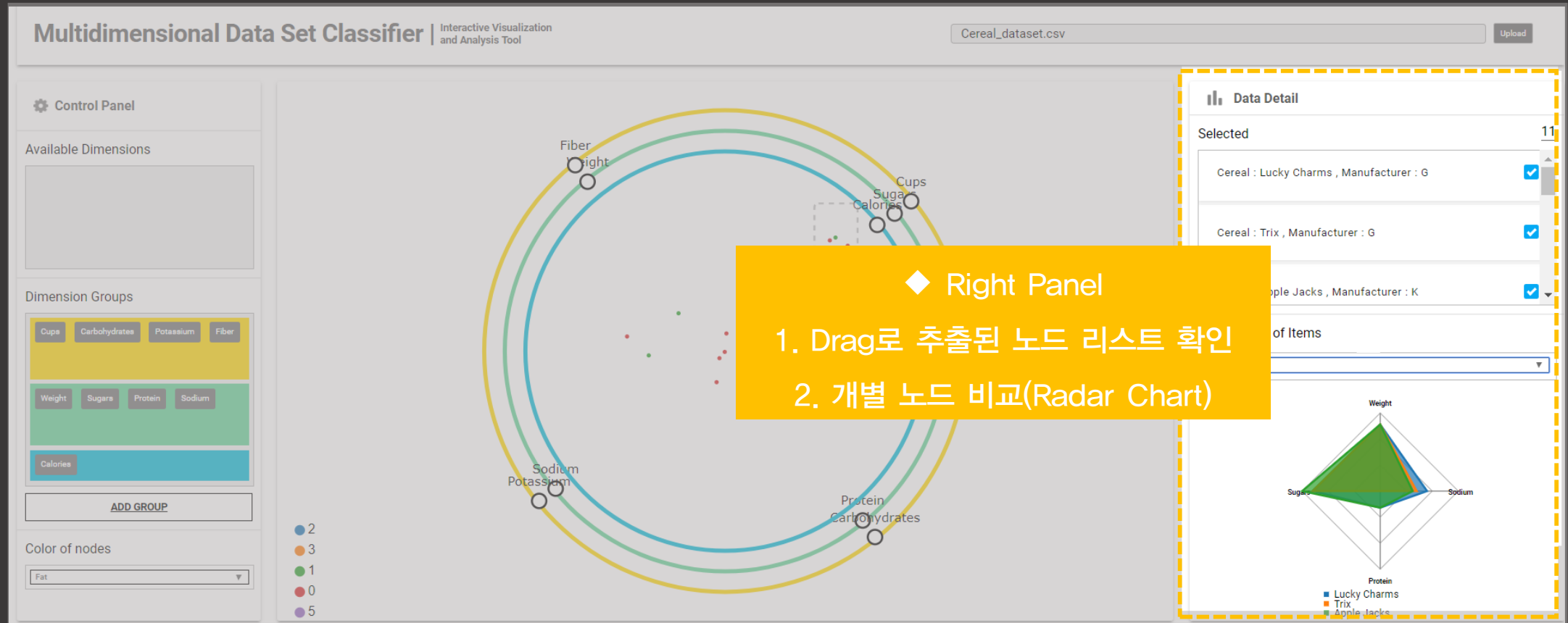
Contents 02.

프로젝트 구현 기능 설명



Contents 02.

프로젝트 구현 기능 설명



Contents 03.

참고 논문

참고 논문 목록

Ono, J. H. P., Sikansi, F., Corrêa, D. C., Paulovich, F. V., Paiva, A., & Nonato, L. G. (2015, August). Concentric RadViz: visual exploration of multi-task classification. In *Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on* (pp. 165–172). IEEE.

Concentric RadViz에 관한 연구, Sigmoid Weighting을 적용한 동심원 방사형 시각화 구현

Gee, A. G., Yu, M., & Grinstein, G. G. (2005). Dynamic and interactive dimensional anchors for spring-based visualizations. *Computer Science, Technical Report University of Massachusetts*

Dimension Anchor의 Spring based Vis에 관한 연구

Nováková, L., & Štěpánková, O. (2006, September). Multidimensional clusters in RadViz. In *Proceedings of the 6th WSEAS International Conference on Simulation, Modelling and Optimization* (pp. 470–475). World Scientific and Engineering Academy and Society (WSEAS).

클러스터링을 하기위한 도구로서의 RadViz에 관한 연구

Daniels, K., Grinstein, G., Russell, A., & Glidden, M. (2012). Properties of normalized radial visualizations. *Information Visualization*, 11(4), 273–300.

Normalized RadViz의 기본적인 특성에 관한 연구

Correlation Analysis on Semiconductor Process Variables Using CCA(Canonical Correlation Analysis) : Focusing on the Relationship between the Voltage Variables and Fail Bit Counts through the Wafer Process

CCA를 통한 반도체 공정 변인들의 상관성 분석에 관한 연구

Contents 04.

차후 계획

앞으로 남은 일



IEEE VIS 2018 Poster 부문

- 6월 16일까지 Submission.
- 선정 시 10월 21일 발표



“데이터 시각화 부문 Paper 기재 및 시각화 툴 웹 서비스”