The background of the entire page is a vibrant green color. Overlaid on this background is a complex, white circuit board pattern. The pattern consists of numerous thin, white lines that represent traces on a PCB, forming a dense, interconnected network. The lines vary in thickness and direction, creating a sense of depth and complexity. The overall effect is that of a high-tech, digital environment.

**SMART DX  
SOLUTION**

**NetMiner 365  
White Paper**

## **EXECUTIVE SUMMARY**

### **1 BACKGROUND & VISION**

- 1.1 Graph
- 1.2 Machine Learning
- 1.3 Graph Machine Learning

### **2 OBJECTIVES**

- 2.1 사용성
- 2.2 독창성
- 2.3 생산성

### **3 SYSTEM COMPONENTS**

- 3.1 Interface
- 3.2 Analysis
- 3.3 Data

### **4 MAIN FEATURES**

- 4.1 Data
- 4.2 Visualization
- 4.3 Analysis

### **5 BENEFITS**

- 5.1 Self-Service Machine Learning
- 5.2 Full-Featured Graph Analytics
- 5.3 Cutting-Edge Graph Machine Learning

### **6 APPLICATIONS**

- 6.1 조직/HR
- 6.2 고객 분석기획
- 6.3 소셜 미디어/언론, 마케팅
- 6.4 지식 그래프/R&D
- 6.5 제조/SCM
- 6.6 범죄/컴플라이언스
- 6.7 바이오

### **7 CONCLUSIONS**

## Executive Summary

NetMiner 365는 누구나 쉽게 그래프 머신러닝 기법을 활용하여 기존과 차별화된 인사이트를 얻을 수 있는 온라인 데이터 과학 플랫폼이다.

빅데이터에는 개체와 그 속성 뿐만 아니라, 행위를 통해 발생한 이벤트(Event)의 내역으로 구성되어 있다. 하지만, 기존 빅데이터 분석이 주로 개체와 그 속성에만 집중한 탓에, 그동안 우리는 빅데이터의 아주 일부분만 이해할 수 있었다. 최근 이러한 한계를 넘어서기 위한 방법으로 '그래프(Graph)'가 주목받고 있다. 그래프는 개체와 개체 간의 연결 혹은 관계로 구조화된 데이터를 의미하며, 그래프 분석(Graph Analytics)은 그래프 구조 상에서 개체의 위치 속성을 계산하는 방법이다. NetMiner 365는 그래프 분석을 최신의 머신러닝 기법과 결합하여, 새로운 관점에서 데이터를 이해할 수 있도록 도와준다.

NetMiner 365는 사용자 친화적인 GUI를 제공하여, 프로그래밍에 익숙하지 않은 사용자도 쉽게 사용할 수 있도록 설계되었다. 또한 그래프 분석, 다변량 통계, 머신러닝 및 딥러닝 등 다양한 분석 기법을 하나의 플랫폼에서 제공함으로써 융복합 분석을 통해 기존과 다른 인사이트를 얻을 수 있다. 그리고 웹을 통해서 언제, 어디에서나 이용할 수 있기 때문에 인프라 구축에 필요한 시간과 비용을 절감할 수 있다.

사용자는 NetMiner 365에서 데이터를 업로드하고, 필요한 대로 전처리 할 수 있다. 또한 데이터를 시각적으로 이해하기 위한 다양한 시각화 알고리즘과 차트를 이용할 수 있다. 그리고 사용자는 몇 번의 클릭으로 간단하게 머신러닝으로 데이터를 학습하고, 학습 모델을 만들어 실제 데이터에 적용하여 데이터를 예측하고 분류할 수 있다. 또한, 그래프에서의 개체의 중요도를 파악하거나, 군집을 파악할 수도 있고, 데이터의 분포량을 파악할 수 있다. 그리고 그래프 머신러닝을 통해 그래프 구조를 학습에 활용하여 예측의 정확도를 향상시킬 수 있다.

NetMiner 365는 HR, 마케팅, 컴플라이언스, R&D 등 다양한 비즈니스 분야에서 활용할 수 있으며, 이미 그 가치를 인정받고 있다. NetMiner 365는 빅데이터 시대에 필요한 기업 가치를 발굴하며, 당면한 과제에 대하여 최적화된 의사결정을 할 수 있도록 도와준다.

# 1. Background & Vision

## 1.1 Graph

그래프(Graph)는 노드(Node, Vertex)와 노드 간의 연결 혹은 링크(Link, Edge) 로 이루어진 데이터 구조이다. 다시 말해, 그래프는 '연결과 관계' 라는 새로운 관점에서 데이터를 이해하므로 기존에 발견할 수 없었던 새로운 인사이트를 제공한다.

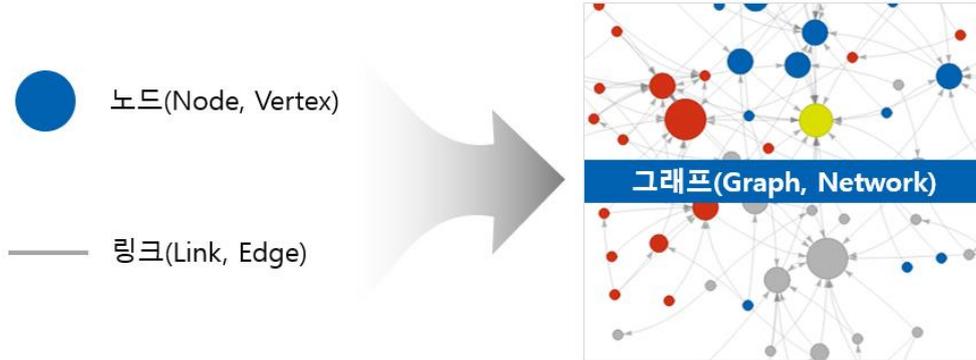


그림 1.1 Graph, Network

특히, Digital Footprint Data인 빅데이터의 상당 부분은 바로 그래프 구조를 가지고 있다. 왜냐하면, 빅데이터는 온라인 상에서 행위자, 개체 간에 발생하는 복잡한 상호작용(Transaction)을 통해 추적 되기 때문이다.

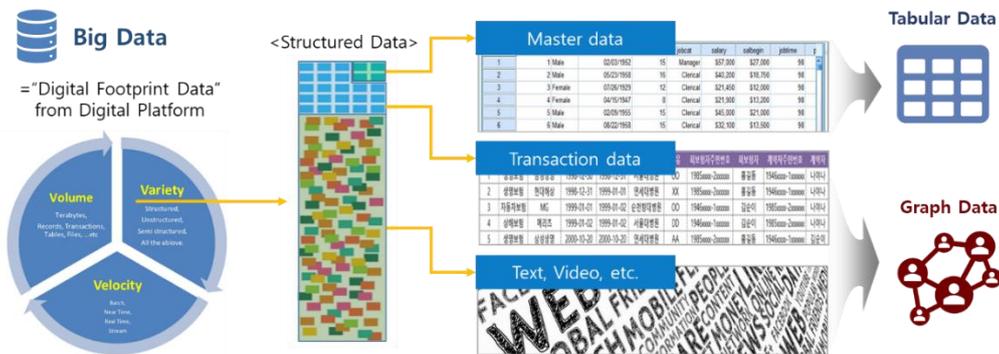


그림 1.2 Digital Footprint Data

예를 들어, 누군가 콘텐츠를 시청하면 시청기록 데이터로 저장되고, 이는 이용자와 콘텐츠와의 관계로 이해할 수 있습니다. 또한 소셜 미디어에서 댓글을 작성하면 이용자 간의 상호 관계로 모델링 할 수 있다.

비정형 데이터도 적절한 전처리를 통해 그래프로 변환할 수 있다. 예를 들어, 텍스트에서 키워드 간의 연관 관계를 구성하여 관심 키워드의 연관 키워드가 무엇인지 파악할 수 있다. 또는 객체 인식 기법을 통해 이미지에서 객체를 뽑아낸 후, 객체 간 그래프로 구성할 수도 있다.

그러나 지금까지 빅데이터 분석은 slice, dice, drill down, roll up, pivot 등의 다차원 분석 (multidimensional analysis)에 국한되어 왔다. 하지만 앞에서 본 것처럼 빅데이터는 기본적으로 그래프 데이터이며, 이러한 그래프 데이터를 제대로 이해하기 위해서는 그에 맞는 분석 방법을 선택하는 것이 중요하다. 그래프 분석(Graph Analytics)은 그래프 데이터에서 구조적 특성과 패턴을 파악하는 고급 분석 방법이며, 기존 빅데이터 분석 방법과 차별화된 새로운 접근 방법으로 주목 받고 있다.

실제로, 글로벌 시장조사기관 Gartner에서도 2019년, 2021년에 그래프 분석을 10대 데이터 분석 기술로 선정하면서 '그래프는 사용자 추천, 머신러닝, 설명 가능한 AI 등의 알고리즘 성능을 향상시킬 수 있는 방법' 이라고 소개한 바 있다.

## 1.2 Machine Learning

머신러닝(기계학습, Machine Learning)은 인공지능의 한 분야로 컴퓨터가 프로그래밍 없이 데이터를 학습하여 새로운 패턴과 지식을 제공하는 방법을 의미한다. 예를 들어, 이전에는 연구자가 직접 데이터를 분류하기 위한 규칙을 정의한 후 규칙을 프로그래밍하여 컴퓨터가 데이터를 분류하게 했다면, 머신러닝에서는 컴퓨터가 데이터를 학습하여 숨겨진 규칙을 찾고 이를 토대로 데이터를 분류하고 예측한다.

머신러닝은 개인화 추천, 이미지 분류, 챗봇, 자동 번역 등 이미 다양한 비즈니스 분야에서 활용되며 그 유용성을 인정받고 있다.

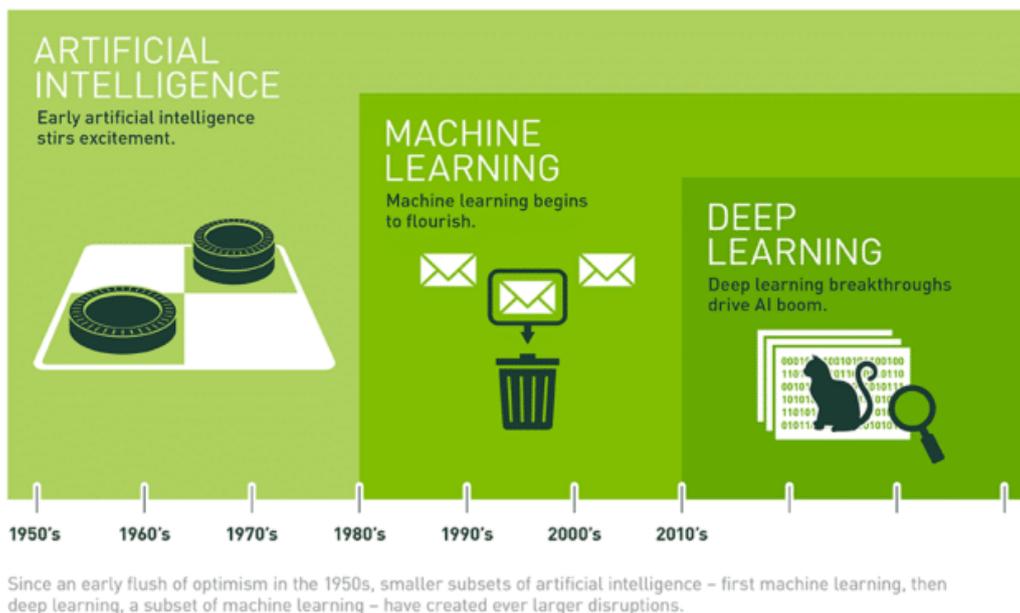


그림 1.3 Machine Learning

\* 출처: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

### 1.3 Graph Machine Learning

그래프 신경망(Graph Neural Network) 은 딥러닝의 일종으로 그래프가 입력 데이터인 신경망을 의미하며, 그래프 머신러닝(Graph Machine Learning)은 그래프에서의 관계적 특성을 머신러닝에 활용하거나 그래프 신경망을 통해 그래프를 학습하는 것을 의미한다. 그래프 머신러닝은 그래프에서의 개체 간 연결 구조를 학습하기 때문에 노드 고유 특성 뿐만 아니라, 주변에 연결된 다른 노드의 특성까지 활용할 수 있다.

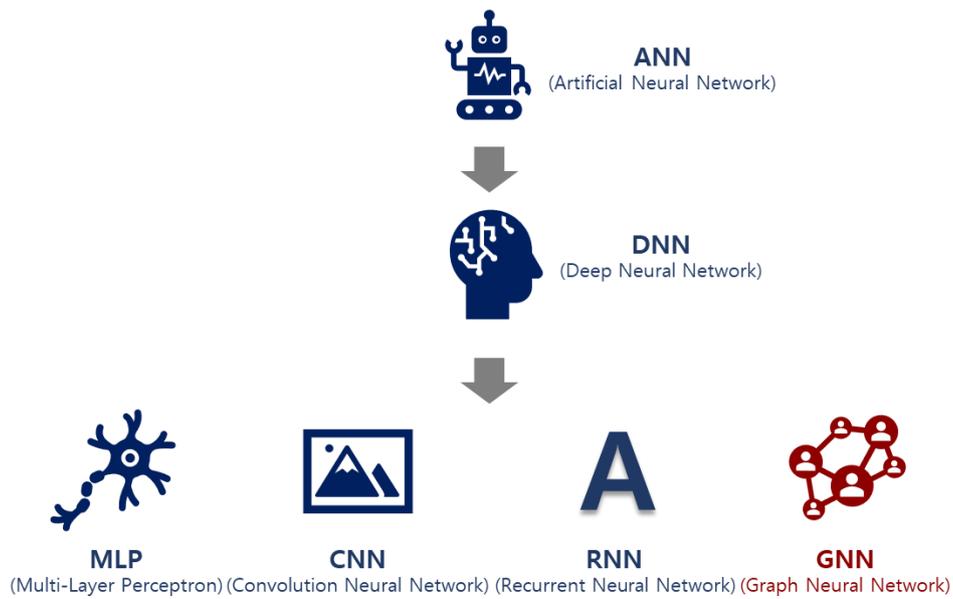


그림 1.4 Graph Machine Learning

그래프 머신러닝은 ‘연결 혹은 관계’ 를 추가로 학습에 활용하기 때문에, 일반적인 머신러닝에 비해 데이터를 더욱 정확하고 정교하게 분류, 예측할 수 있다. 그래프 머신러닝은, 다음과 같은 기능을 제공한다.

- ① 노드 분류(Node Classification): 노드의 특성과 연결 관계를 기반으로 노드를 분류할 수 있다. 대표적인 응용 사례로는 온라인 소셜미디어에서의 이용자 성향 분류, 인용 네트워크를 활용한 논문 주제 분류 등이 있다.
- ② 링크 예측(Link Prediction): 서로 다른 개체 사이의 링크 및 연관성을 예측하는 문제입니다. 대표적인 예로 스트리밍 서비스에서의 콘텐츠 추천, 온라인 소셜미디어에서의 친구 추천, 좋아할 만한 사진 추천(Pinterest) 등이 있다.
- ③ 그래프 분류(Graph Classification): 그래프 자체를 카테고리별로 분류하는 문제입니다. 예를 들어, 조직의 소통 구조에 따른 유형 분류, 단백질 분자 구조 기반의 물질 분류 등을 위해 활용될 수 있다.

## 2. Objectives

NetMiner 365를 통해 이용자는 이전에는 얻을 수 없었던 새로운 효용을 경험할 수 있다. 특히 인력 리소스, 컴퓨팅 환경, 개발 프레임 워크 등의 문제로 쌓여 있는 빅데이터를 제대로 활용할 수 없었다면, NetMiner 365는 매우 유용한 도구가 될 것이다.

### 2.1 사용성

NetMiner 365는 사용자 친화적인 GUI 기반의 데이터 과학 플랫폼으로 몇 번의 클릭만으로 쉽고 간편하게 데이터를 업로드하고 분석할 수 있다. NetMiner 365의 메뉴 구조는 데이터 처리 (Data), 시각화 (Visualize), 분석 (Analysis) 과 같이 일반적인 데이터 분석의 흐름에 따라 설계되었다.

또한 프로그래밍 언어를 사용하지 않아도 직관적으로 데이터를 다루고 분석할 수 있도록 표준화된 프로세스에 따라 제작되어 쉽게 사용 방법을 익힐 수 있다. 예를 들어, NetMiner 365의 데이터 분석 기능은 '분석 목적 → 데이터 선택 → 알고리즘 선택'의 단계로 표준화되어 쉽게 사용 방법을 익힐 수 있다. 게다가 단계 별로 선택 가능한 항목만 표시해주거나, 알고리즘의 파라미터들을 자동으로 제시해주기 때문에 NetMiner 365에 익숙하지 않은 사용자도 제시되는 항목을 선택하기만 하면 충분히 NetMiner 365를 이용할 수 있다.

### 2.2 독창성

NetMiner 365는 그래프와 머신러닝이 결합된 세계 유일의 데이터 과학 플랫폼으로, 사용자는 NetMiner 365로 매우 독창적이고 차별화된 결과를 얻을 수 있다. 기존의 빅데이터 분석은 통계, 조건 검색 등 주로 개체 고유의 특성(Feature)에 집중하는 방법이었다. 하지만, '관계'를 분석하는 그래프를 활용하면, 기존에는 확인할 수 없었던 새로운 정보를 획득할 수 있고 데이터 예측의 정확도를 높일 수 있다.

먼저, 그래프는 빅데이터를 이해할 수 있는 새로운 패러다임으로, 이미 여러 비즈니스 분야에서 그래프를 활용하고 있다. 대표적으로 Google은 웹 문서 간 하이퍼링크 관계를 그래프로 구성하고 PageRank라는 그래프 분석 알고리즘을 적용하여, 웹페이지의 상대적 중요성을 평가함으로써 검색 결과의 질을 크게 향상시켜서 최고의 검색 엔진이 될 수 있었다. 또한, Amazon은 판매 데이터에서 고객과 상품 간 그래프를 구성하여 그래프 내에서의 등위성 개념을 응용한 협업 필터링 (Collaborative Filtering) 알고리즘을 상품 추천에 활용하기도 하였다. NetMiner 365는 20년 간 사이람(주)이 쌓아온 그래프/네트워크 분석의 노하우를 바탕으로 설계 되어 그래프 분석에 필수적인 기능들을 제공한다.

또한, 머신러닝과 딥러닝은 최신의 데이터 분석 방법으로 이미 많은 분야에서 유용성을 인정받고 있다. NetMiner 365는 일반적으로 널리 사용되는 데이터 분류와 회귀를 위한 머신러닝 및 딥러닝

알고리즘을 제공한다. 사용자는 간단하게 샘플 데이터를 학습시키고 학습 모델을 실제 데이터에 적용하여 결과를 예측하거나 데이터를 분류할 수 있다. 예를 들어, 의료 분야에서는 기존 환자의 건강 정보(심박수, 혈당, 키, 체중 등)와 병명을 머신러닝을 활용하여 학습하고, 이후 새로운 환자의 건강 정보를 넣어주면 예상 병명을 제시할 수 있다.

마지막으로 NetMiner 365는 그래프 머신러닝을 위한 다양한 파이프라인을 제공한다. NetMiner 365는 그래프 분석의 결과를 머신러닝의 Feature로 이용할 수도 있고, 그래프 자체를 학습할 수 있는 그래프 신경망 알고리즘(GNN)도 제공한다.

## 2.3 생산성

NetMiner 365는 클라우드 기반의 온라인 데이터 과학 플랫폼이기 때문에, 언제, 어디에서나 웹을 통해 액세스가 가능하다. 또한 시스템 및 하드웨어 제약이 줄어들어 별도의 인프라 구축 없이도 쉽게 빅데이터 분석을 시작할 수 있다.

사용자의 작업 공간에 데이터를 저장해 두므로, 데이터 분석과 시각화를 위해서 매번 데이터를 불러오지 않아도 된다. 또한, 하나의 데이터에 여러 개의 알고리즘을 동시에 적용할 수 있기 때문에, 다양한 결과를 빠르고 쉽게 비교할 수 있다.

그리고 분석 결과를 다시 새 데이터로 저장할 수 있어서, 다른 분석 알고리즘에서 활용하거나 시각화 결과에 표현할 수 있다. 예를 들어, 그래프에서 측정한 중심성(Centrality) 점수를 그래프 시각화에서 '점'의 크기에 반영하면, 가장 중요한 노드가 무엇인지 시각적으로 알 수 있다.

### 3. System Components

NetMiner 365는 온라인 데이터 과학 플랫폼으로 인터페이스, 분석, 데이터, 인프라로 구성되어 있다.

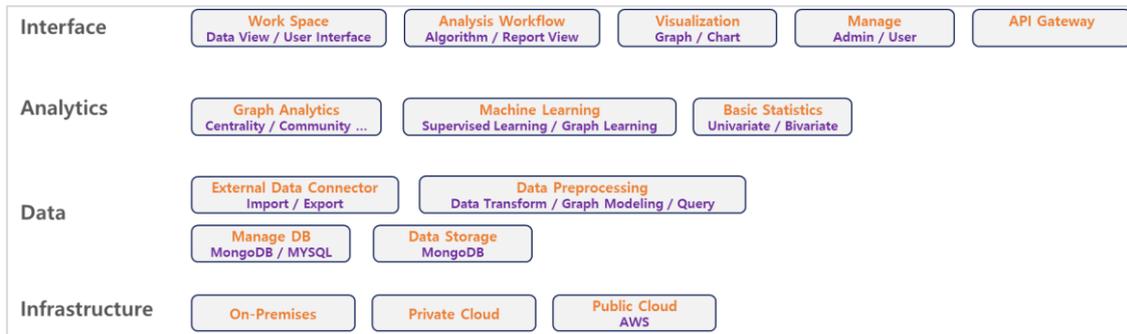


그림 3.1 System Components

#### 3.1 Interface

인터페이스는 워크스페이스 (Workspace), 분석 워크플로우 (Analysis Workflow), 시각화 (Visualization), 관리 (Manage), API 게이트웨이 (API Gateway) 로 구분된다.

##### 3.1.1. 워크스페이스

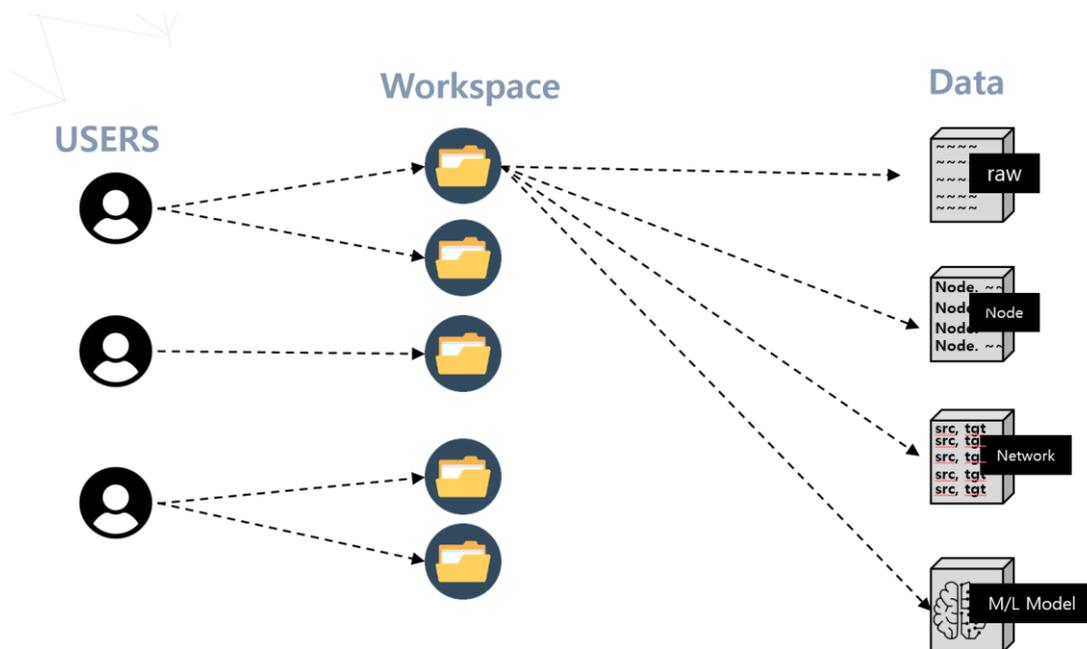


그림 3.2 워크스페이스

개별 사용자가 데이터 분석을 위한 독립되고 한정된 자원의 공간을 워크스페이스라고 하며, 워크스페이스에서는 사용자가 데이터를 확인(Data View)하고 관리하고, 데이터 시각화 및 분석을 서버에 요청할 수 있다.

### 3.1.2. 분석 워크플로우

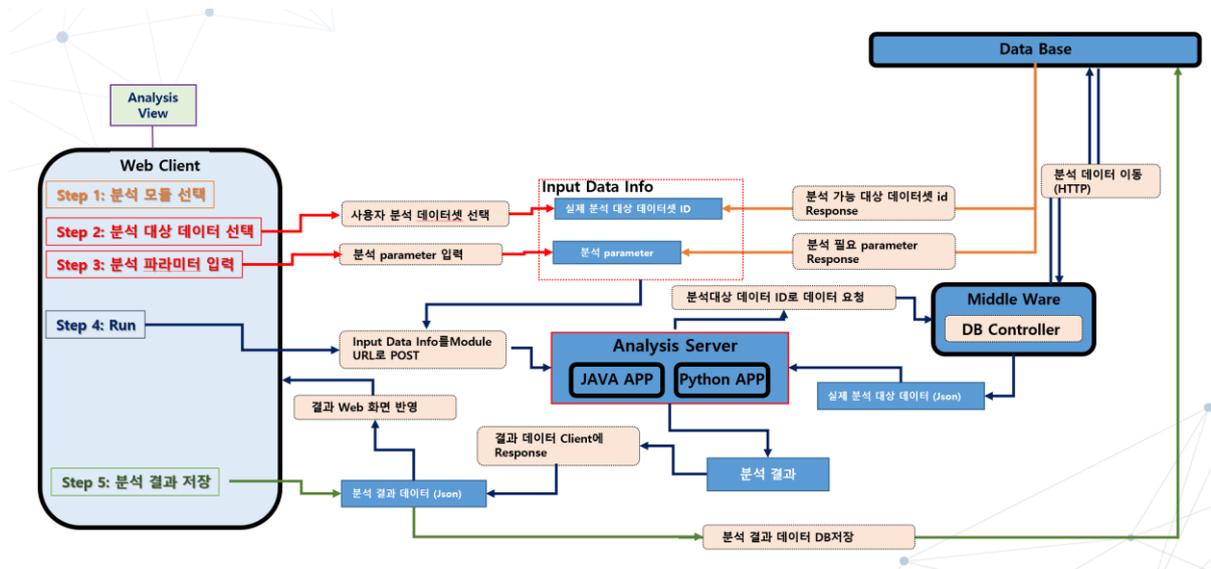


그림 3.3 분석 워크플로우

선택한 입력 데이터의 유형에 따라 다양한 종류의 분석을 선택할 수 있으며, 분석 목적에 맞는 알고리즘을 선택하고, 알고리즘의 파라미터 조정이 가능하다.

### 3.1.3. 시각화

그래프 데이터를 기반으로 한 그래프 시각화(Graph Visualization)과 데이터의 분포를 시각적으로 확인할 수 있는 차트 시각화(Chart Visualization)로 구분된다.

그래프 시각화는 그래프의 노드와 링크를 점과 선으로 표현하여 전체적인 그래프 구조를 시각적으로 보여준다. 또한, 노드와 링크의 특성을 점과 선에 반영하는 스타일링 기능을 제공한다.

차트 시각화는 그래프 데이터 뿐만 아니라, Tabular 데이터 (행과 열로 구성된 테이블 데이터)에 대해 다종의 차트(Bar chart, Line chart, Histogram, Pie chart, Tree map, Boxplot, Scatter Plot, Parallel Coordinate, Rader chart, Matrix Diagram)로 시각화 할 수 있는 기능이다.

### 3.1.4 관리

사용자는 자신의 계정에 할당된 가용 자원과 분석 수행 로그를 확인할 수 있다. 또한, 서비스 운영을 위해 모든 사용자의 가용 자원과 분석 수행 로그를 확인할 수도 있다.

### 3.1.5 API 게이트웨이

사용자는 자신의 계정에 할당된 가용 자원과 분석 수행 로그를 확인할 수 있다. 또한, 서비스 운영을 위해 모든 사용자의 가용 자원과 분석 수행 로그를 확인할 수도 있다.

## 3.2 Analytics

Data Storage에 업로드 한 데이터를 기반으로 분석을 수행하며, 모든 알고리즘은 각각 개별 파라미터를 설정할 수 있으며, 고유의 분석 결과를 산출한다.

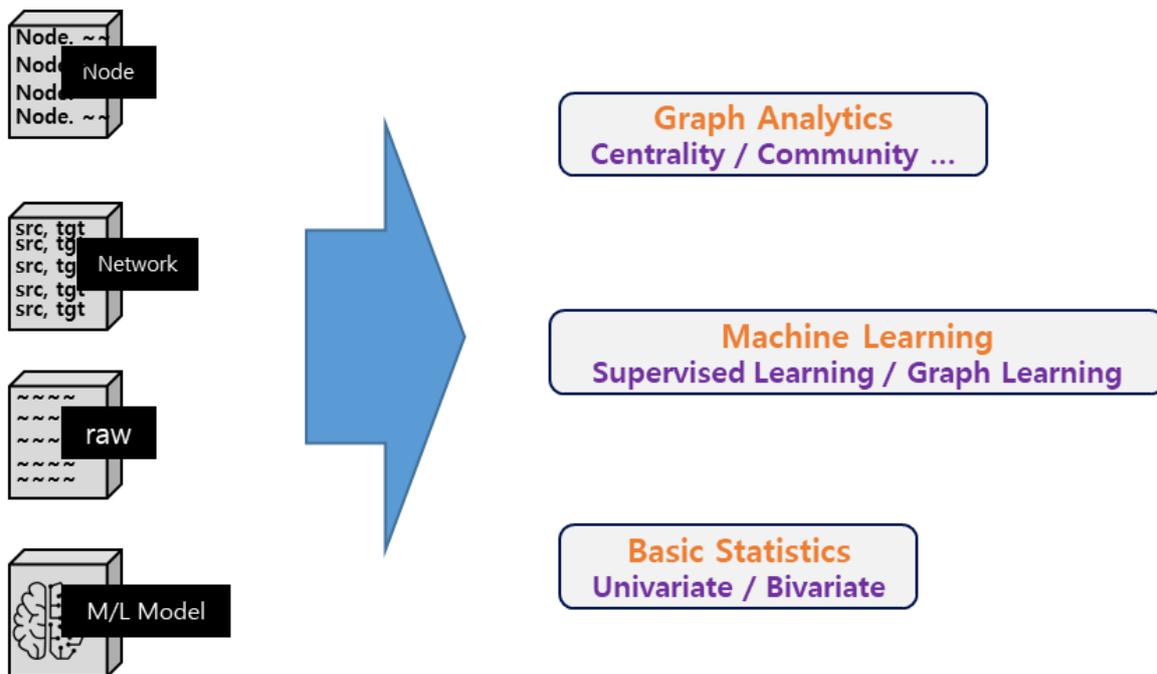


그림 3.4 애널리틱스

### 3.2.1. 그래프 분석 (Graph Analytics)

그래프 데이터를 대상으로 분석을 수행합니다. 총 5개의 분석 유형(Centrality, Community, Equivalence Cluster, Node Embedding, Graph Properties)을 제공하며, 각 분석 유형에는 여러 개의 세부 알고리즘이 포함되어 있다. JAVA Application으로 개발되었다.

### 3.2.2 머신러닝 Machine Learning)

총 3개의 분석 유형(Supervised Learning, Unsupervised Learning, Graph Learning)을 제공한다. 각 분석 유형에 따라 선택할 수 있는 데이터가 달라지는데, Supervised Learning과 Unsupervised Learning은 일반적인 Tabular 데이터 혹은 노드 데이터를 대상으로 한다. 반면, Graph Learning은 그래프 데이터를 대상으로 한다. Supervised Learning의 경우에는 데이터를 학습하여 모델을 생성하는데, 모델을 저장하면 이후에 Deploy 할 수 있다. Python Application으로 개발되었다.

### 3.2.3 기초 통계 Basic Statistics)

총 2개의 분석 유형(Univariate, Bivariate)을 제공하며, 단일 변수 및 다중 변수 통계분석 알고리즘을 포함하고 있다. JAVA Application으로 개발되었다.

## 3.3 Data

데이터 수집, 가공, 처리 등의 모든 부분을 담당하고 있다.

### 3.3.1 Manage DB

사용자의 가입 및 라이선스 정보와 라이선스 및 가용 자원에 대한 다양한 정보들을 포함한다. 가입정보는 MySQL에서 담당하며, MySQL의 라이선스 정보를 MongoDB 라이선스와 연결하여 사용자별 데이터가 연동된다.

### 3.3.2 Data Storage

실질적으로 사용자의 데이터가 저장되는 공간이며, Manage 엔진을 기반으로 사용자의 데이터 가용 정보가 관리된다. 사용자는 데이터 스토리지에 저장된 데이터만 분석 대상으로 선택할 수 있으며, 데이터는 세 가지 유형(Tabular, Node, Network)으로 구분된다. 추가로 기계학습 분석에 의해 학습된 모델 데이터와 시각화 정보를 포함한 스타일링 정보도 여기에 저장된다.

### 3.3.3 External Data Connector

외부 DB (MySQL, Oracle, MSSQL) 및 파일(csv, xlsx)을 가져오거나, 사이람㈜ 소셜 네트워크 분석 소프트웨어인 NetMiner 4에서 저장한 파일 (nmf)을 가져올 수 있다. 저장된 데이터는 파일(csv, excel)로 내보낼 수 있으며, 분석 결과 데이터도 역시 내보내기가 가능하다.

### 3.3.4 Data Processing

사용자의 필요에 따라 데이터를 전처리 하고 가공하는 기능을 포함하고 있다. Tabular 데이터에서 그래프 데이터를 추출할 수 있는 그래프 모델링 (Graph modeling) 기능, 데이터 쿼리, 그래프 데이터에 대한 다양한 전처리 알고리즘을 포함한다.

## 4. Main Features

NetMiner 365는 ① Data (데이터 처리) ② Visualization (데이터 시각화) ③ Analysis (데이터 분석)를 위한 기능을 제공한다.

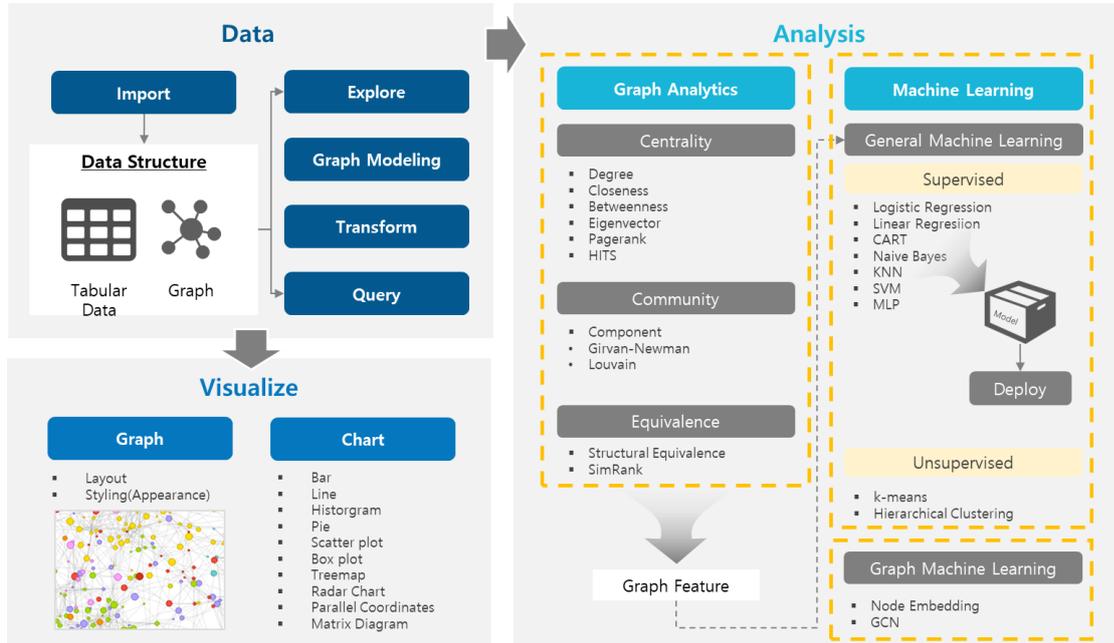


그림 4.1 Main Features

### 4.1 Data

데이터를 가져오고, 가져온 데이터를 전처리 하는 기능이다. 특히, NetMiner 365는 다음과 같이 일반적인 Tabular 데이터뿐만 아니라 그래프 데이터를 위한 데이터 구조를 가지고 있다.

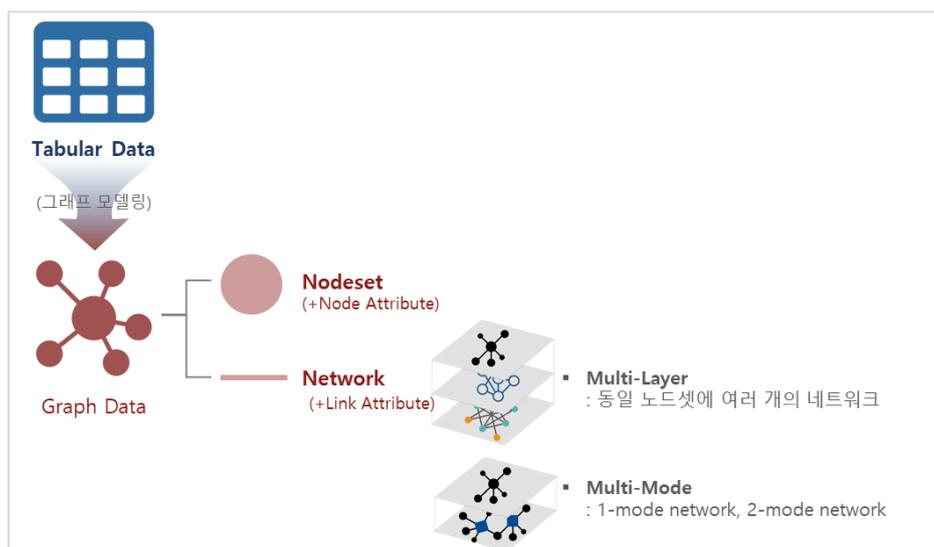


그림 4.2 데이터

### 4.1.1 Import

외부 데이터를 NetMiner 365에 가져오는 기능이다. 파일(csv, xlsx)과 DBMS(MySQL, Oracle, Microsoft SQL server)에서 데이터를 가져올 수 있다.

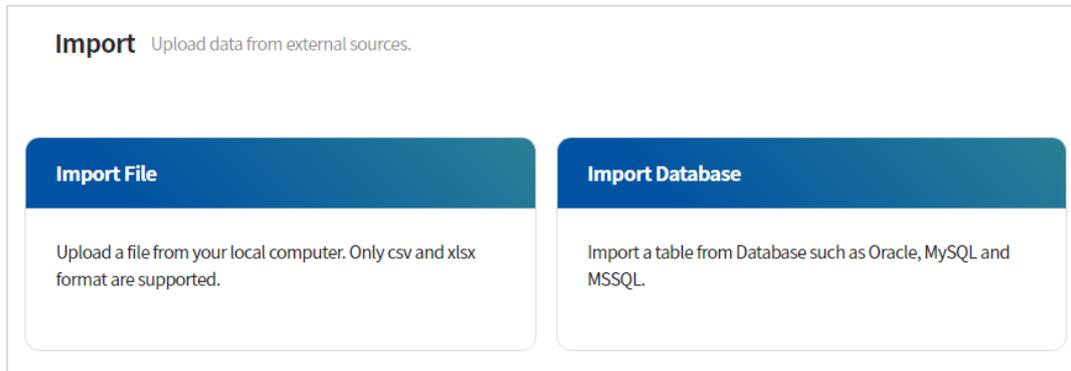


그림 4.3 Import

### 4.1.2 Explore

Import를 통해 NetMiner 365에 업로드한 데이터를 확인하고 (View), 데이터 삭제, 복사, 내보내기 등을 할 수 있다.

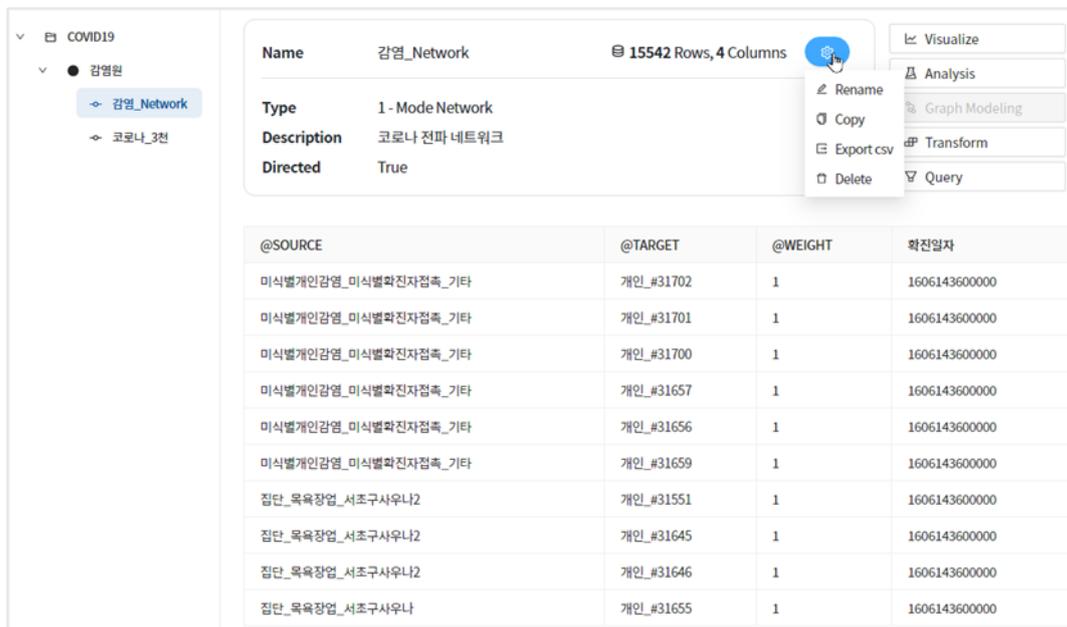


그림 4.4 Explore

### 4.1.3 Graph Modeling

Tabular 데이터에서 그래프 구조를 정의하여 그래프 데이터로 추출합니다. 예를 들어, 판매 로그에서 고객과 상품 간의 그래프 관계를 추출할 수 있다.

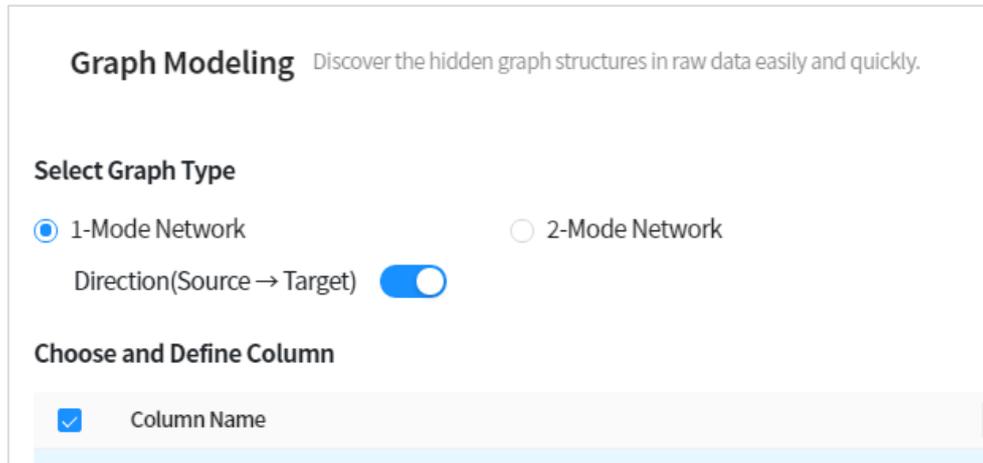


그림 4.5 Graph Modeling

### 4.1.4 Transform

저장한 데이터를 필요에 따라 변환할 수 있다. 예를 들어, 데이터 값을 로그 처리하거나, 그래프의 방향을 바꾸는 등의 변환 알고리즘을 제공한다. 변환된 데이터는 새 데이터 아이템으로 저장할 수 있다.

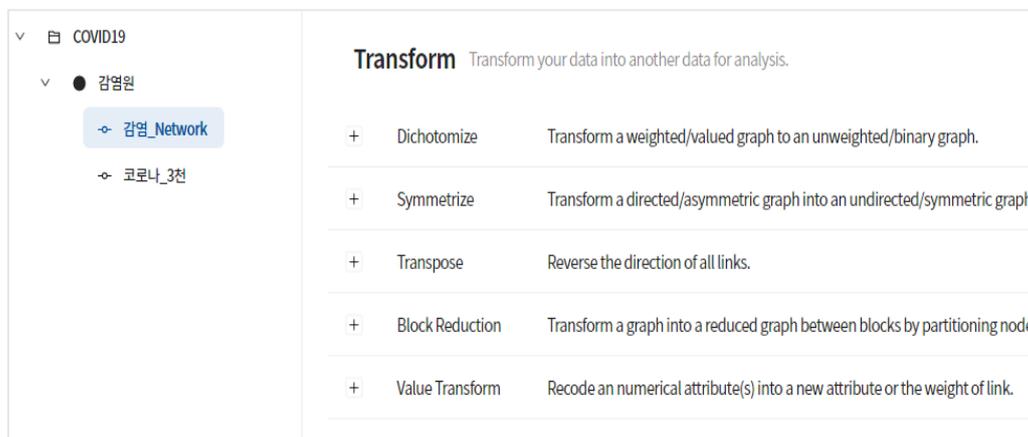


그림 4.6 Transform

### 4.1.5 Query

조건을 주어 데이터를 추출할 수 있다. 예를 들어, 특정 시점 이후에 데이터만 추출하거나, 특정한 속성을 가진 데이터만 추출하여 새 데이터 아이템으로 저장할 수 있다.

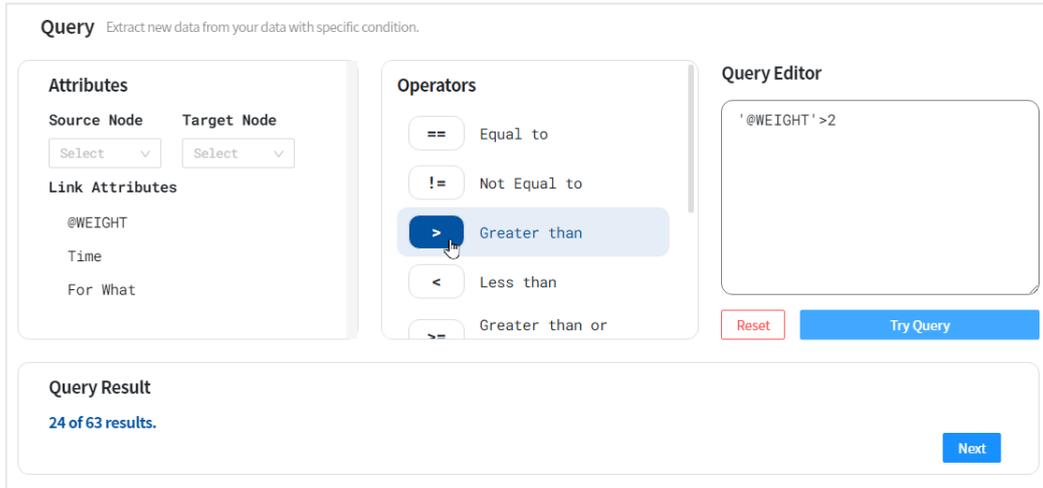


그림 4.7 Query

## 4.2 Visualize

데이터를 시각적으로 표현하기 위한 기능이다. NetMiner 365는 일반적인 차트 뿐만 아니라 그래프 데이터를 위한 시각화 기능도 제공한다.

### 4.2.1 Graph Visualization

노드를 점으로, 노드 간 연결 관계를 선으로 시각화하여 그래프 구조를 통합된 화면에서 볼 수 있게 도와준다. 또한 노드 및 링크의 고유 속성 혹은 데이터 분석 결과를 점과 선의 스타일에 반영함으로써 다양한 정보를 한 장의 이미지에 담을 수 있다.

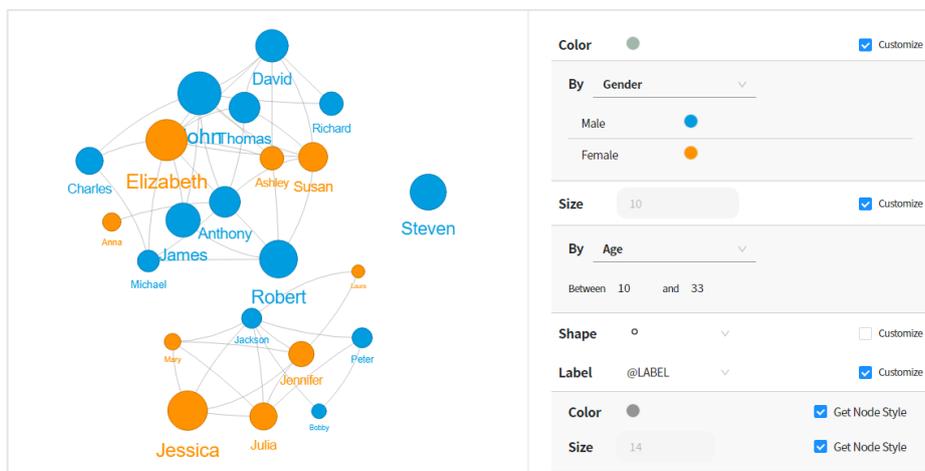


그림 4.8 Graph Visualization

## 4.2.2 Chart

데이터를 시각적으로 탐색할 수 있도록 다양한 차트를 제공한다. 참고로 데이터 간 연결 관계를 시각화하여 그래프 모델링에 참고할 수 있는 평행 좌표(Parallel Coordinates)도 제공한다.

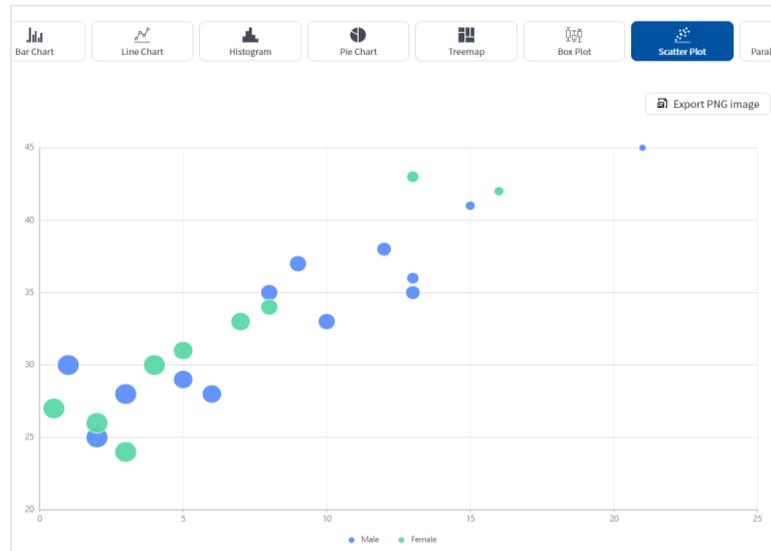


그림 4.9 Chart

## 4.3 Analysis

NetMiner 365는 다양한 데이터를 다각도로 이해할 수 있도록 세 가지 데이터 분석 방법을 제공한다. 하나의 플랫폼에서 여러 데이터 분석 방법을 융합함으로써, 기존과는 다른 새로운 인사이트를 얻을 수 있다.

### 4.3.1 Graph Analytics

그래프 데이터를 이해하기 위한 필수적인 알고리즘을 제공한다. 사용자는 그래프 분석을 통해 다음과 같은 결과를 얻을 수 있다.

- ① **Centrality (중심도):** 그래프에서 노드의 중요성 및 영향력을 측정합니다. 어떤 노드가 핵심 플레이어인지, 주변부에 존재하는 노드가 무엇인지를 확인할 수 있다.
- ② **Community (커뮤니티):** 서로 촘촘히 연결되어 있는 응집 클러스터를 탐지한다 (Community Detection). 노드 속성의 유사성이 아니라 연결 구조를 기반으로 데이터를 클러스터링 할 수 있다.
- ③ **Equivalence Cluster (등위 클러스터):** 노드의 역할 및 지위를 판별한다.

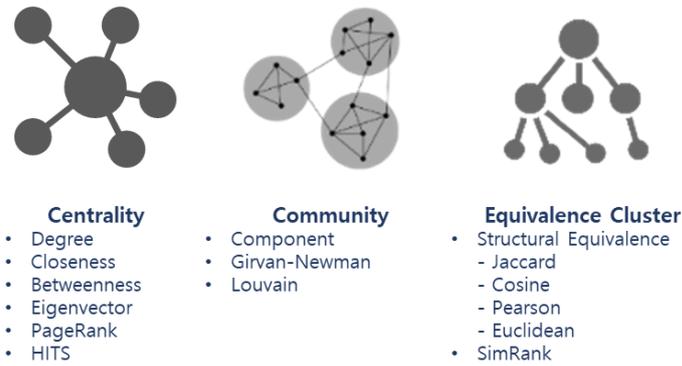


그림 4.10 Graph Analytics

### 4.3.2 General Machine Learning

범용적으로 널리 사용되는 머신러닝 및 딥러닝 알고리즘을 제공한다. 사용자는 머신러닝을 통해 샘플 데이터를 학습시켜서 학습 모델을 만들 수 있다. 이렇게 만들어진 학습 모델을 실제 데이터에 적용하면, 쉽게 데이터를 분류하고 예측할 수 있다. 또한 데이터 분포 및 상관 관계를 파악하기 위한 다변량 통계 알고리즘도 함께 제공한다.

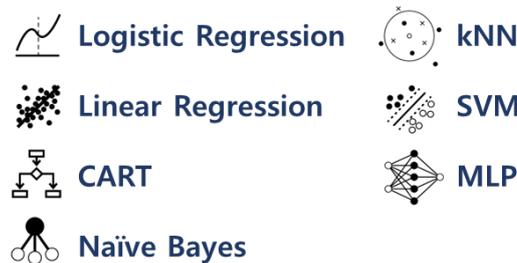


그림 4.11 General Machine Learning

### 4.3.3 Graph Machine Learning

그래프 데이터를 학습할 수 있는 알고리즘과 파이프라인을 제공합니다. 그래프 머신러닝은 노드의 특성만을 이용하는 일반 머신러닝과는 달리, 노드가 가지고 있는 연결 관계도 함께 활용하기 때문에 더욱 정확히 데이터를 예측하고 분류할 수 있다.

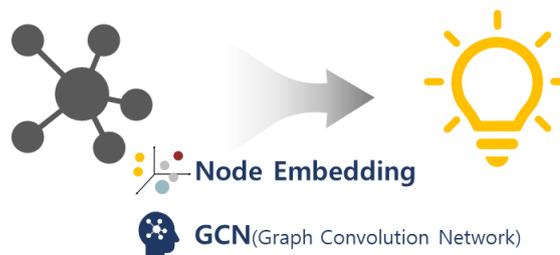


그림 4.12 Graph Machine Learning

## 5. Benefits

NetMiner 365는 이용자에게 다음과 같은 세 가지 이점을 제공한다.

- ① Self-Service Machine Learning
- ② Full-Featured Graph Analytics
- ③ Cutting-Edge Graph Machine Learning

### 5.1 Self-Service Machine Learning

NetMiner 365를 통해 이용자는 몇 번의 클릭만으로 간단하게 머신러닝을 할 수 있다. NetMiner 365는 GUI 기반의 데이터 과학 플랫폼으로 프로그래밍에 익숙하지 않은 이용자도 쉽게 이용할 수 있도록 설계되었다.

#### 5.1.1 지도 학습

이용자는 단 3단계만에 데이터 예측 및 분류 모델을 만들 수 있다. ① 학습을 위한 샘플 데이터를 선택하고 ② 학습할 Feature를 선택하고 ③ 머신러닝 알고리즘을 선택하기만 하면 된다.

예를 들어, A 스트리밍 서비스 구독 데이터가 있다고 가정하겠다. NetMiner 365에서 이용자는 고객의 서비스 구독 여부를 연령 정보와 함께 머신러닝 알고리즘에 학습시킬 수 있다. 그러면, 머신러닝 알고리즘이 고객의 이용 현황과 개인정보와의 연관성을 파악하여, 나이에 따른 서비스 구독 여부를 예측할 수 있는 학습 모델을 생성한다.

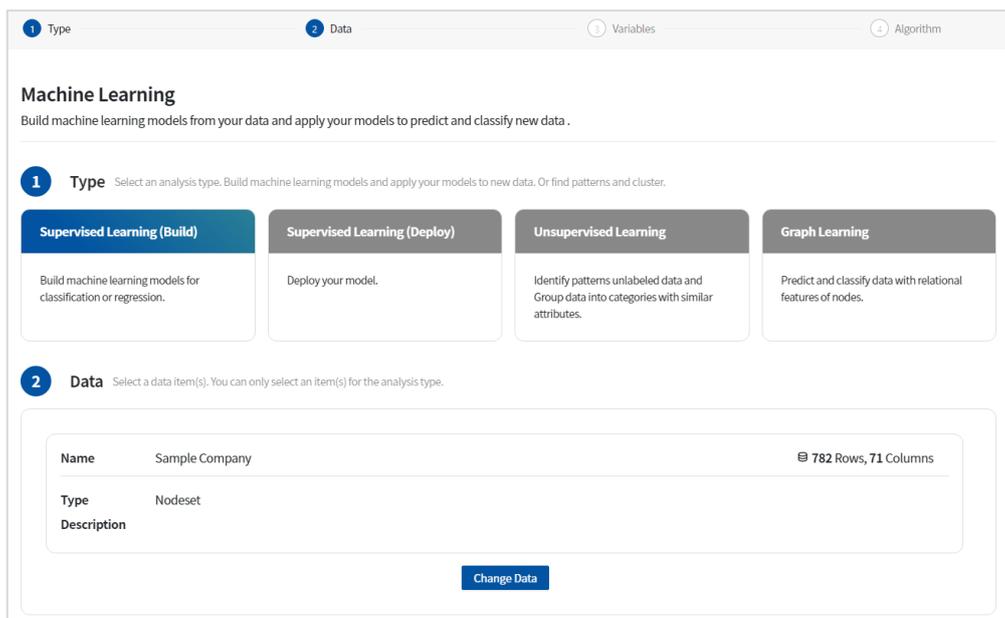


그림 5.1 지도 학습

특히, NetMiner 365에서는 하나의 데이터에 여러 머신러닝 알고리즘을 한 번에 적용할 수 있어서 만들어진 모델의 성능을 한 눈에 비교하고 가장 좋은 성능의 모델을 골라 저장할 수 있다. 저장한 학습 모델은 이후 실제 데이터를 예측하고 분류하는데 활용할 수 있다.

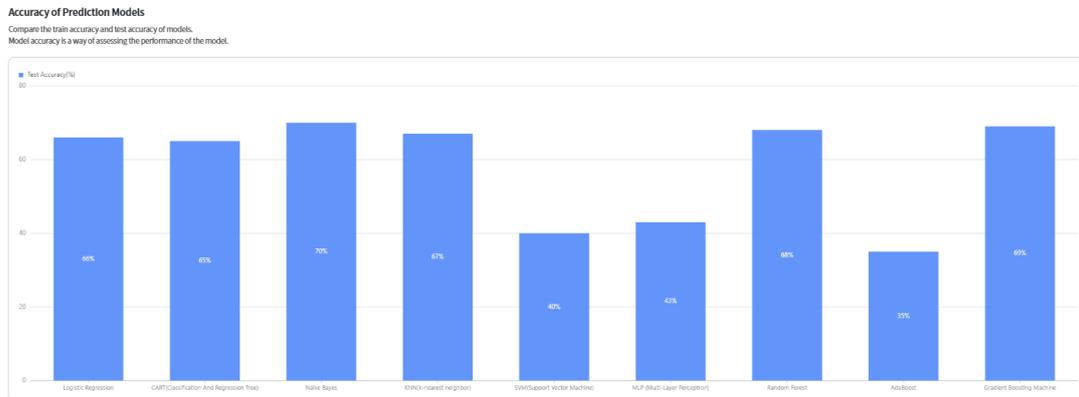


그림 5.2 머신러닝 알고리즘 비교

### 5.1.2 비지도 학습

Feature의 유사 정도에 따라 데이터를 분류 및 클러스터링 할 수 있다.

예를 들어, 고객의 서비스 이용 빈도, 성별, 나이, 지역 등의 데이터를 토대로 유사한 특성을 가진 고객 군집(클러스터)을 발굴할 수 있다. 이를 통해 고객의 관심을 유도할 수 있는 적합한 콘텐츠를 제작할 수 있고, 맞춤형 마케팅이 가능하다.

## 5.2 Full-Featured Graph Analytics

NetMiner 365는 그래프 분석을 위해 필요한 모든 것을 제공한다.

### 5.2.1 그래프 데이터 구조 및 전처리 기능

NetMiner 365의 데이터 구조는 그래프의 구성 요소인 노드 데이터와 네트워크 데이터를 저장할 수 있도록 설계되었다. 노드 데이터는 여러 개의 네트워크 데이터를 가질 수 있어서, 동일한 노드 사이에 존재하는 다양한 관계를 표현할 수 있다. 그리고 노드와 링크의 속성도 함께 저장할 수 있어서 데이터 시각화와 분석에 유용하게 사용할 수 있다.

또한 그래프 분석을 위한 다양한 전처리 기능도 제공하고 있어서, 필요에 따라 데이터를 가공할 수 있다.

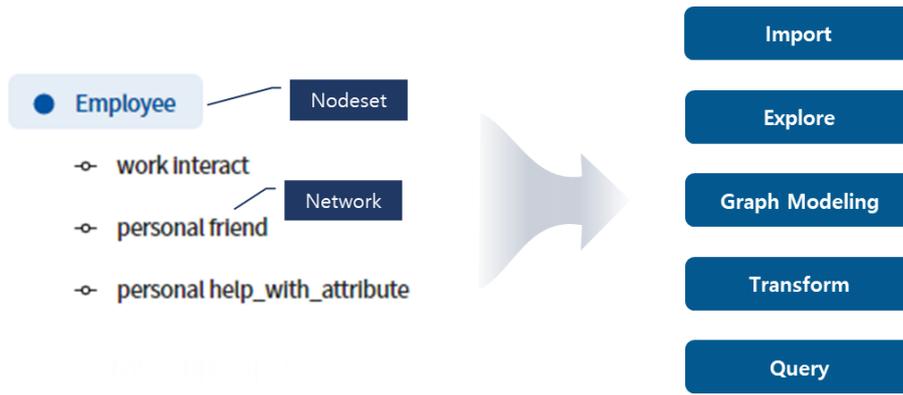


그림 5.3 그래프 데이터 구조 및 전처리 기능

### 5.2.2 그래프 시각화

그래프 구조를 시각적으로 표현하기 위한 레이아웃 알고리즘을 제공한다. 노드는 점, 링크는 선으로 표현하며, '관계'를 기반으로 노드의 위치를 계산하여 연결된 노드 간은 가깝게, 연결이 없는 노드는 멀리 배치한다. 그러므로, 이용자는 한 눈에 서로 밀접하게 연결되어 있는 노드들은 무엇인지, 다른 노드와 연결이 희박한 노드들은 무엇인지 파악할 수 있다.

또한 이용자는 스타일링 기능을 통해 그래프 시각화에 다양한 정보를 담을 수 있다. 예를 들어, 키워드 연관 그래프에서 키워드의 사용 빈도가 높을수록 점의 크기를 크게 스타일링 하면, 많이 언급된 중요한 키워드가 무엇인지 강조해서 표현할 수 있다.

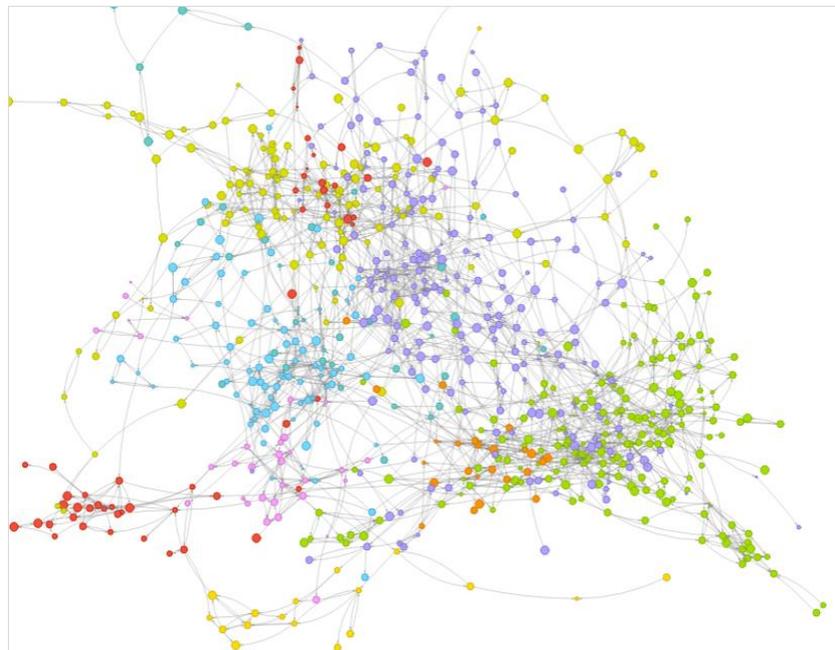


그림 5.4 그래프 시각화

### 5.2.3 그래프 분석

그래프 데이터를 이해하기 위한 그래프 분석 알고리즘을 제공한다.

그래프에서의 노드의 중요성을 측정하기 위한 Centrality 알고리즘, 연결이 밀집된 커뮤니티를 탐지하는 Community 알고리즘, 노드의 역할 및 위치를 파악하여 분류할 수 있는 Node Similarity (Equivalence Cluster) 알고리즘을 제공한다.

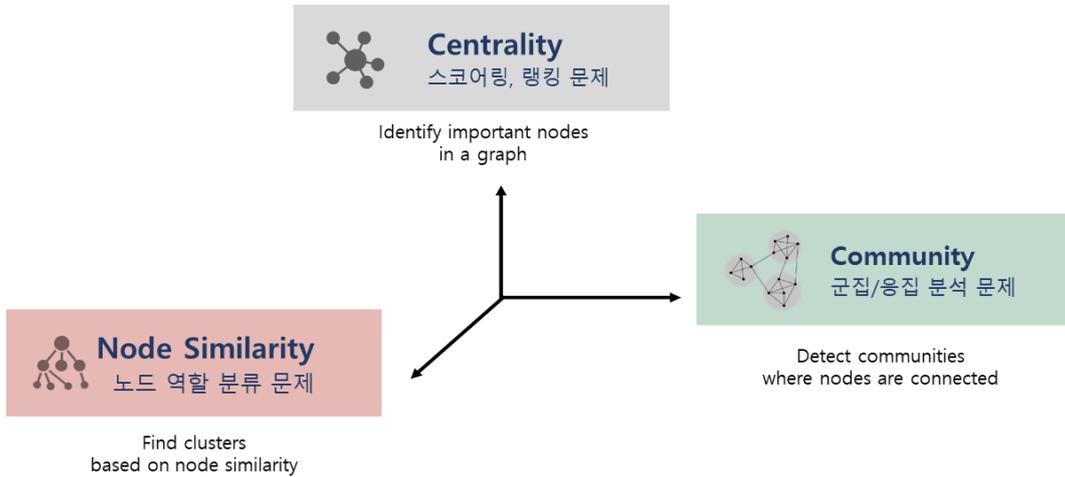


그림 5.5 그래프 분석

## 5.3 Cutting-Edge Graph Machine Learning

NetMiner 365는 다양한 그래프 머신러닝 파이프라인을 지원한다.

### 5.3.1 그래프 분석 결과를 일반 머신러닝에 활용

그래프 분석 알고리즘을 활용하여 Centrality, Community, Node Similarity 결과를 저장하면, 이를 일반적인 머신러닝 알고리즘의 Feature 로 사용할 수 있다.

예를 들어, 온라인 소셜 미디어 이용자의 정치 성향을 판별하는 문제가 있다. 정치 성향은 이용자의 성별, 나이, 지역, 사용 단어 등으로 파악할 수 있다. 추가로 이용자가 팔로우 (Follow) 하는 다른 이용자의 성향도 정치 성향을 파악하는데 도움이 될 수 있다. 그러므로, 머신러닝을 통해 이용자 분류 모델을 만들 때, 이용자의 기초 정보 뿐만 아니라, Follow 관계에서의 커뮤니티 판별 결과를 사용하면 더욱 정확하게 성향을 판별할 수 있다.

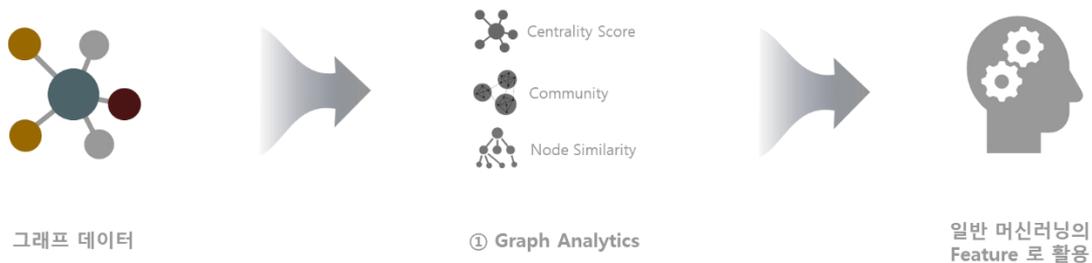


그림 5.6 그래프 분석 결과를 일반 머신러닝에 활용

### 5.3.2 그래프 구조를 임베딩(Node Embedding)하여 일반 머신러닝에 활용

그래프 구조를 2차원 벡터 상에 임베딩(Embedding)하면, 그래프 구조를 머신러닝에서 학습의 Feature로 사용할 수 있다. 노드 임베딩(Node Embedding)은 컴퓨터가 그래프 데이터를 이해할 수 있도록 숫자의 나열인 벡터 형태로 바꾸는 것을 의미한다.

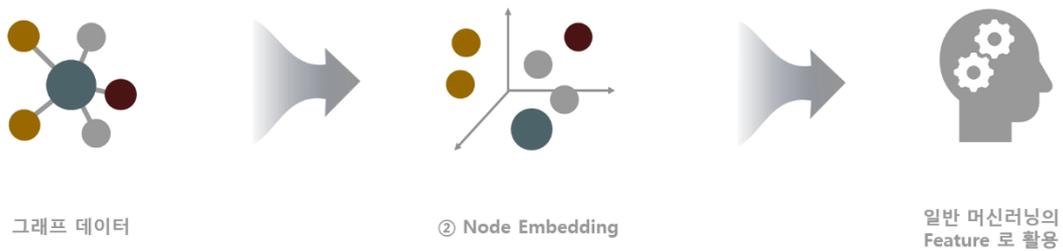


그림 5.7 그래프 구조를 임베딩(Node Embedding)하여 일반 머신러닝에 활용

### 5.3.3 그래프 신경망을 통한 그래프 머신러닝

그래프 신경망은 그래프 구조와 노드의 Feature를 바로 학습하여 데이터를 분류하고 예측할 수 있다. 그래프 신경망은 최근 딥러닝 분야에서 부상하고 있는 최신의 데이터 분석 방법으로, NetMiner 365는 GCN(Graph Convolution Network)라는 알고리즘을 제공하고 있다.

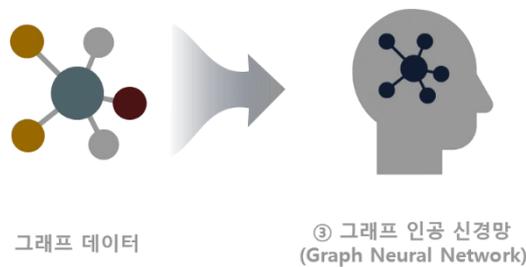


그림 5.8 그래프 신경망을 통한 그래프 머신러닝

## 6. Applications

### 6.1 조직 / HR

조직 시스템을 구성하는 지식, 업무, 자원, 사람들끼리 발생하는 다양한 층위의 네트워크는 조직의 성과와 발전에 매우 중요한 요소이다. 조직에서 발생한 실제 커뮤니케이션 패턴을 수집하여 소통의 구조와 효율성을 진단함으로써 기업은 조직의 흐름을 더 잘 이해하고 조직의 성과를 효과적으로 개선할 수 있다.

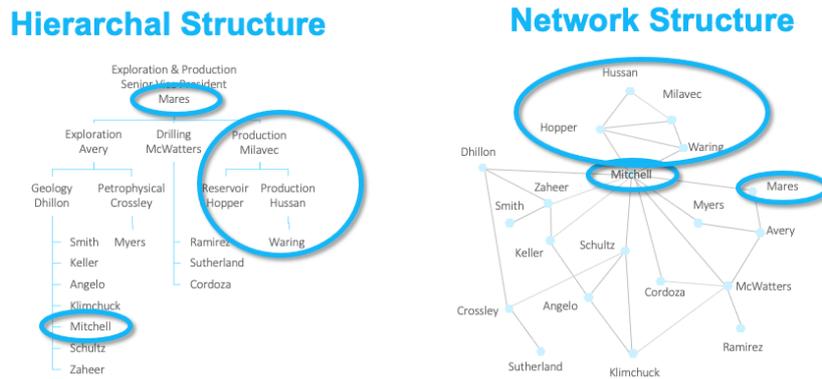


그림 6.1 조직도와 협업 그래프

\* 출처: <https://www.i4cp.com/productivity-blog/what-organizational-network-analysis-is-and-how-it-benefits-companies>

### 6.2 고객 분석 / 기획

새로운 소비의 시대. 고객의 행위 패턴과 관심사를 파악할 수 있는 수많은 데이터가 존재한다. 고객과 이벤트 데이터, 고객 간의 상호 작용 등을 그래프로 구조화하고, 그래프 분석과 머신러닝 분석을 통해 고객의 행위와 확산 패턴을 분석할 수 있다. 이를 통해 마케터는 고객의 프로필과 행동을 예측하여 개인 맞춤형 제안을 할 수 있다.

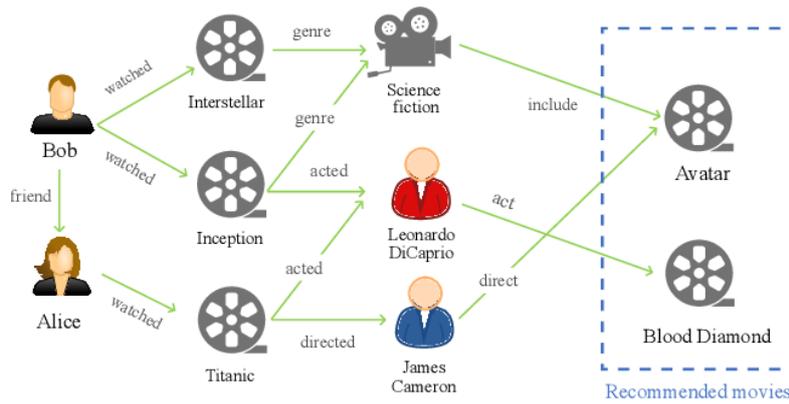


그림 6.2 콘텐츠 추천

\* 출처: <https://www.catalyzex.com/paper/arxiv:2003.00911>

## 사례

핀터레스트는 이미지 공유를 기반으로 한 소셜 미디어 서비스이다. 핀터레스트는 이용자에게 더욱 정교하게 이미지를 추천하기 위해서 그래프 머신러닝 알고리즘인 PinSage를 시스템에 적용했다. 기존의 이미지 추천 방식은 이용자가 저장한 이미지와 모양과 색상이 유사한 이미지는 추천해 주었지만 비슷한 관심사를 가진 다른 이용자의 취향은 반영할 수 없었다. 하지만, 이용자와 이미지라는 '관계'를 활용하면서 더욱 다양하고 정확하게 이미지를 추천할 수 있게 되었다.

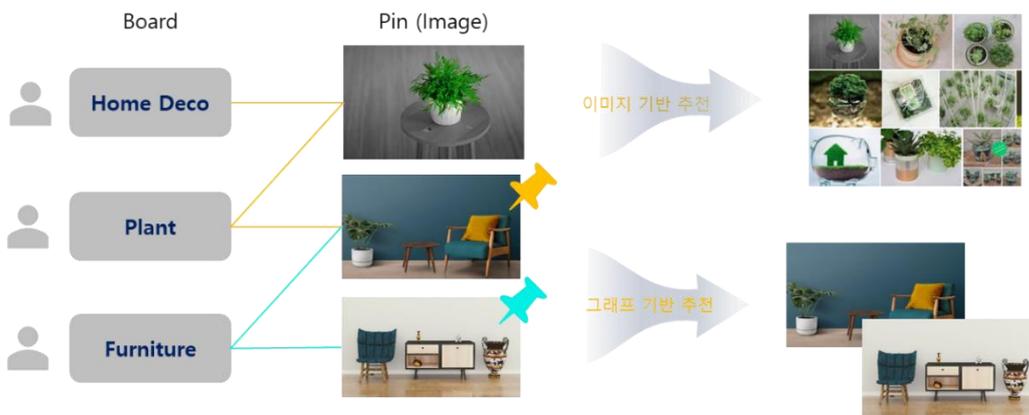


그림 6.3 핀터레스트 사례

## 6.3 소셜 미디어 / 언론, 마케팅

소셜 미디어의 이슈에 대한 온라인 여론을 파악하고, 트렌드를 이해할 수 있다. 그리고 소셜 미디어 사용자들의 정보 전파 및 확산 영향력을 측정하여 인플루언서를 파악하거나, 행동 패턴 및 성향에 따라 이용자를 분류할 수 있다.

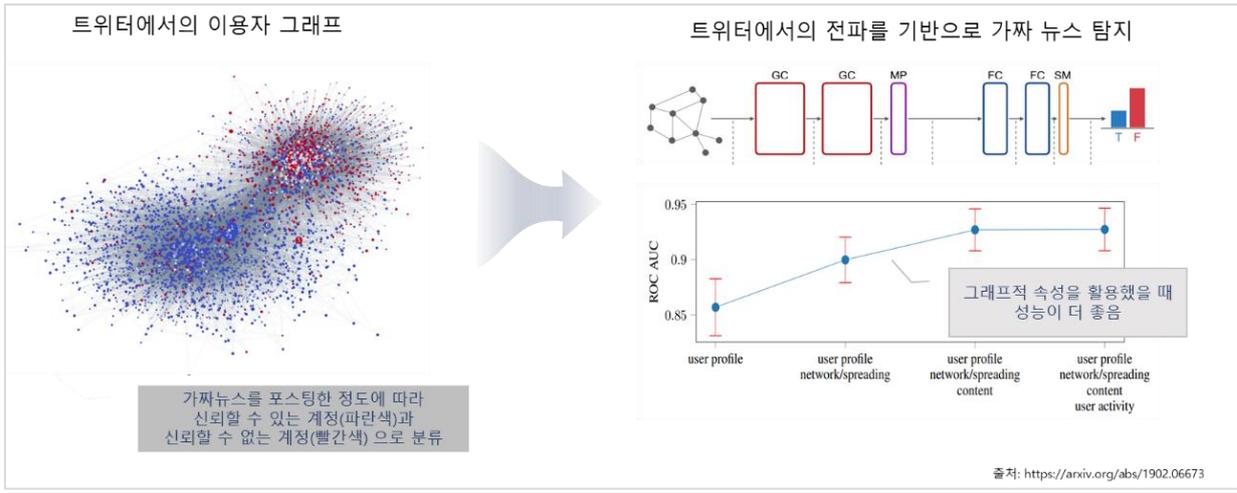


그림 6.4 트위터 사용자 분류

### 6.4 지식 그래프 / R&D

그래프를 기반으로 한 지식 지도는 방대한 분량의 문헌 데이터에 존재하는 지식의 구조를 시각적으로 이해할 수 있도록 도와주며, 머신러닝은 데이터의 분류 및 예측에 드는 시간과 비용을 크게 줄여줄 수 있다. 이를 통해 기술 트렌드를 이해하고 전문가를 쉽게 파악할 수 있다.

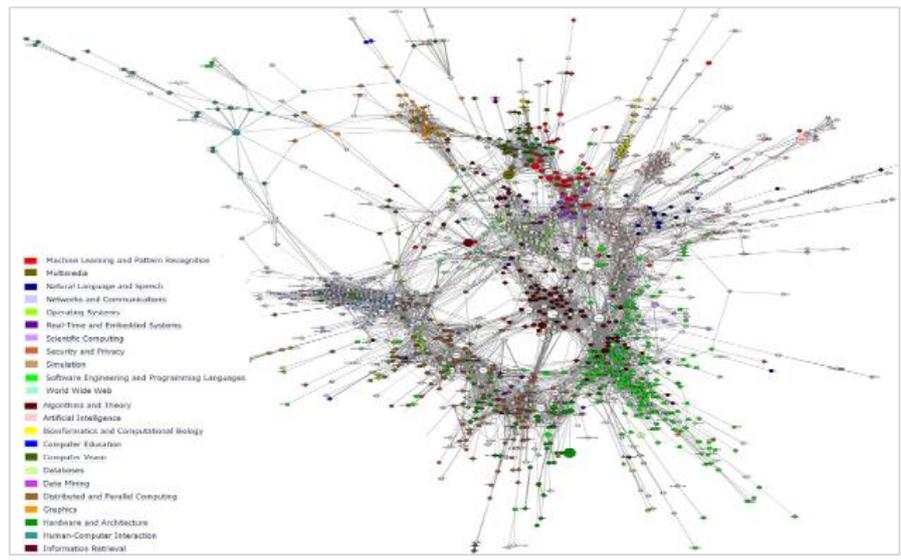


그림 6.5 컴퓨터 사이언스 분야의 지식 그래프

\* 출처: [https://www.researchgate.net/figure/The-computer-science-knowledge-network-giant-component\\_fig2\\_221273592](https://www.researchgate.net/figure/The-computer-science-knowledge-network-giant-component_fig2_221273592)

### 6.5 제조 / SCM

생산 과정을 효율화 하고, 제품의 수요와 공급량을 정확하게 예측하고, 물류 경로나 조건을 최적

화하면 기업의 수익을 크게 증가시킬 수 있다. 그래프 분석은 공정 구조, 다른 기업과의 거래 관계, 물류 이동 패턴 등을 파악하여 손실을 줄이고, 공급망의 병목을 해소하는 것을 도울 수 있다.

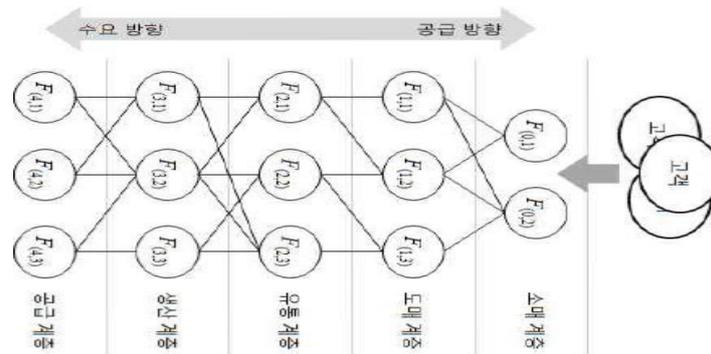


그림 6.6 공급 사슬 그래프 구성

\* 출처: <https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=TRKO201300010302&dbt=TRKO>

## 6.6 범죄 / 컴플라이언스

사회 시스템 고도화로 인해, 사기, 공모, 부정행위의 패턴도 점차 복잡해지면서 추적에 많은 시간과 비용이 든다. 그래프 분석과 머신러닝을 활용하면 빠르게 이상 패턴을 파악하고, 전체 혐의 그룹의 관계 구조를 이해할 수 있다.

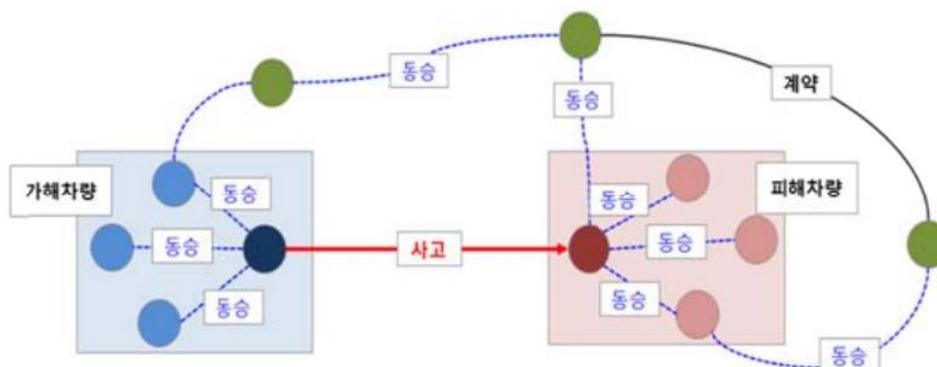


그림 6.7 자동차 손해 보험 공모 사기 적발 사례 (금융감독원)

\* 출처: [http://www.cyram.com/img/download/solution\\_fraud.pdf](http://www.cyram.com/img/download/solution_fraud.pdf)

## 6.7 바이오

바이러스 전파 및 감염 경로를 그래프로 구성하면, 전파 및 확산의 경로와 속도를 예측할 수 있다. 또한 특정 질병, 증상, 약물 성분 간의 관계를 그래프로 구조화해서, 성분 조합을 통해 새로운 신약 개발에 활용할 수 있다.

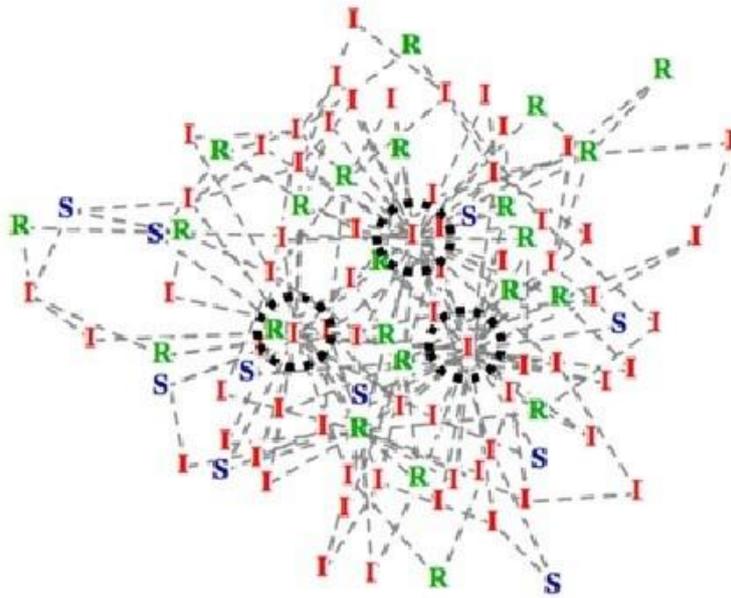


그림 6.8 전염병 확산 그래프

\* 출처: [https://www.hani.co.kr/arti/science/science\\_general/927522.html](https://www.hani.co.kr/arti/science/science_general/927522.html)

## Conclusion

인터넷, 모바일, 5G 등 엄청난 IT 기술 발달 덕분에 우리는 바야흐로 빅데이터 시대를 살아가고 있다. 하지만, 한 설문조사에 따르면, 국내 중소기업의 80%가 사업에 빅데이터를 활용한 적이 없다고 응답<sup>1</sup>할 만큼 기업에서 빅데이터를 제대로 활용하지 못하고 있다. 빅데이터 분석 기술을 도입할 때 가장 큰 걸림돌은 인프라 구축 및 도입에 필요한 비용과 인력 확보를 꼽았다.

NetMiner 365는 누구나 쉽게 데이터를 분석하고 시각화 할 수 있도록 설계되었기 때문에 그동안 기업에서 가지고 있던 이슈를 해결할 수 있다. NetMiner 365의 사용자 친화적인 인터페이스 덕분에 기업은 프로그래머, 개발자, 데이터 과학자 등의 추가적인 인력을 확보할 필요가 없다. 특히, NetMiner 365에는 데이터를 이해하는데 필요한 다양한 데이터 시각화 및 분석 기법들이 탑재되어 있기 때문에, 사용자가 직접 복잡한 개발 언어나 라이브러리를 배울 필요가 없다.

또한 클라우드 기반의 SaaS (Software as a Service) 이기 때문에 고성능의 컴퓨팅 환경을 별도로 구축하지 않고도 데이터 분석을 할 수 있다. 물론, 기업의 보안을 위해서 프라이빗 클라우드에도 설치할 수 있도록 구현되었다.

그리고 NetMiner 365는 '그래프'라는 기존과 다른 관점으로 빅데이터를 분석할 수 있기 때문에, 사용자는 다른 데이터 분석 솔루션 및 플랫폼과 차별화되고, 더욱 정교한 분석 결과를 얻을 수 있다. 특히, 그래프는 최근 기존의 데이터 분석 한계를 극복하기 위한 방법으로 부상하고 있으며,

<sup>1</sup> 출처: <https://www.mk.co.kr/news/economy/view/2019/08/654130/>

구글, 아마존, 핀터레스트 등 여러 기업에서 그래프 분석을 비즈니스에 도입하고 있다.

결론적으로, 기업은 적은 비용으로 NetMiner 365를 도입하여 데이터 분석 결과를 비즈니스에 바로 활용하거나, 또는 대규모의 빅데이터 분석 시스템 도입을 위한 사전 테스트 베드로서 활용할 수도 있다.

---



나무기술은 가상화부터 클라우드, 인프라, 빅데이터, AI 에 이르기까지  
고객의 디지털 혁신과 효과적인 운영을 리딩하는 클라우드 솔루션 기업입니다.

나무기술(주) : 서울시 강서구 마곡중앙8로3길 41 | 02-3288-7900

사이람 : 경기도 성남시 분당구 대왕판교로 670 U-Space B 동, 904 호 | T 031-739-8352

이메일 : [sales@namutech.co.kr](mailto:sales@namutech.co.kr)

홈페이지 : [www.namutech.co.kr](http://www.namutech.co.kr)