|  |
| --- |
| **빅데이터 분석 정의서** |

|  |  |
| --- | --- |
| **주제 :** | 머신러닝을 이용한 전통주 추천 알고리즘 |

|  |
| --- |
| **2023.08.17** |

**Ⅰ. 개요**

1. 아이디어 주제

: 머신러닝을 이용한 전통주 추천 알고리즘

2. 개발 목표

: 서비스 이용자의 연령, 선호 맛, 가격 등 입력 정보에 따른 전통주 추천

3. 개발 내용

: 사용자 입력 데이터와 전통주의 데이터 중 가장 유사도가 높은 전통주 추천

: 설문조사 각 데이터를 전통주 데이터 중 가장 근접한 전통주 찾기

: 사용자 입력 데이터와 설문조사 데이터 중 가장 유사도가 높은 전통주 추천

**Ⅱ. 기능별 빅데이터 분석 명세서**

|  |  |
| --- | --- |
| **기능명** | **酒BTI** |
| 1. **데이터 준비** | |
| **데이터 정의** | 각 전통주에 대한 맛, 가격 등 정보  예상 사용자의 전통주 선호도 |
| **데이터 획득 방법** | 1. 한국농수산식품유통공사\_전통주 정보   https://www.data.go.kr/data/15048755/fileData.do   1. 술담화 사이트 크롤링   https://www.sooldamhwa.com/   1. 설문조사 |
| **수집 데이터** | 전통주 206개에 대한 맛(단맛, 신맛, 바디감, 탄산), 주종, 가격, 도수, 어울리는 안주, 보관방법, 맛 설명  설문을 통한 사용대상자의 성별, 연령과 전통주 선호 맛, 가격, 주종 |
| **2. 전처리** | |
| **전처리 과정** | 중복된 술 데이터 삭제  누락된 술 정보 데이터 추가  단위통일  맛, 도수, 가격 MinMaxScaler 적용  주종, 연령, 성별 필터링  설문조사의 성별, 연령, 주종 수치화 |
| **전처리 데이터** |  |
| **3. 데이터 분석** | |
| **데이터 분석 목표** | 사용자 입력에 따른 전통주 추천 |
| **데이터 분석 시나리오** | : 사용자 입력 데이터와 전통주의 데이터 중 가장 유사도가 높은 전통주 추천  : 설문조사 각 데이터를 전통주 데이터 중 가장 근접한 전통주 찾기  : 사용자 입력 데이터와 설문조사 데이터 중 가장 유사도가 높은 전통주 추천 |
| **데이터 분석 결과** | 결과를 표시 |
| **4. 모델 생성 및 학습** | |
| **모델링 목표** | 사용자가 원하는 맛, 가격, 도수 등과 가장 비슷한 전통주 추천 |
| **학습 모델** | KNN  사용자 기반 협업필터링  아이템 기반 협업 필터링 |
| **학습 방법** | 사용자의 성별과 연령과 선호하는 맛, 도수, 가격 입력 |
| **모델 예측 결과** |  |
| **5. 검증** | |
| **분석 검증 방안** | 구현한 모델에 대한 검증 방안 나열 |
| **분석 평가 결과** | 학습 후 모델에 대한 평가 결과 |