**데이터 분석 계획서**

|  |  |  |
| --- | --- | --- |
| **구분** | **내용** | **비고** |
| 팀명 | Crime BigData Team = CBT. |  |
| 팀원(역할) | OOO – 데이터 분석을 위한 R 및 Jupyter  OOO – 가시화(웹) 및 RDBMS 연동  OOO – DB 테이블구축, SQL 및 DB 모델링  OOO – HADOOP, SPARK, 크롤링 |  |
| 기간 | 2018.8.12. ~ 2018.8.28 (8.30 서울 발표) |  |
| 제목 | Crime System for Visualization = CSV. |  |
| 분석주제 | 범죄와 관련된 정보를 이용하여 범죄위험지역를 선정하고 해결방안을 제시할 수 있는 시스템 |  |
| 분석목적 | 범죄에 영향을 미치는 요인을 분석하여 개선방안 모색 |  |
| 분석배경 | 강진 여고생 실종∙살인 사건, 제주 여성 실종 사건 등이 이슈화되고 있는 요즈음, 범죄에 영향을 미치는 요인들을 고려하여 이러한 사태를 사전에 대비할 수 있는 방법을 찾고자 함 |  |
| 분석대상 | 전국 범죄 현황을 다양한 요인들을 통해 심층 분석 |  |
| 가설설정 | 여러 사회적 요인이 범죄에 영향을 줄 것이다. |  |
| 분석데이터 | 통계청 및 데이터 프리존  -범죄 종류별: 범죄발생시간, 발생도시, 연령 등의 데이터  -유동 인구 및 상권 데이터  웹크롤링을 통한 기사 자료 활용 |  |
| 분석방법 | -수집된 자료를 바탕으로 관련성이 높은 범죄 요인 정리  -대용량 데이터 처리(하둡, 스파크)  -R을 이용하여 Correlation Analysis 및 시각화 |  |
| 분석결과활용 | 범죄의 발생 확률이 높은 지역에 주요 요인들을 보강하여 범죄율을 줄일 수 있는 솔루션 제공 |  |
| SWOT 분석 | |  |  |  | | --- | --- | --- | |  | Opportunity[기회] | Threat[위협] | | Strength[강점] | 내용  .  .  . | 내용  .  .  . | |  | SO 전략 | ST 전략 | |  | WO 전략 | WT 전략 | | Weakness[약점] | 내용  .  .  . | 내용  .  .  . | |  |

**시스템 구축 계획서 양식**

|  |  |  |
| --- | --- | --- |
| **구분** | **내용** | **비고** |
| 팀명 | Crime BigData Team = CBT. |  |
| 팀원(역할) | OOO – 데이터 분석을 위한 R 및 Jupyter  OOO – 가시화(웹) 및 RDBMS 연동  OOO – DB 테이블구축, SQL 및 DB 모델링  OOO – HADOOP, SPARK, 크롤링 |  |
| 기간 | 2018.8.12. ~ 2018.8.28 (8.30 서울 발표) |  |
| 시스템 명 | Crime System for Visualization = CSV. |  |
| 구축 내용 | 범죄와 관련된 정보를 이용하여 미래 범죄에 대응할 수 있도록 범죄위험지역를 선정하고 해결방안을 제시할 수 있는  시스템 |  |
| feature(특징) | * 범죄 관련 데이터 간의 연관성 분석 * 지역별 특성과 범죄의 관계 분석 * 특정 상업들과 범죄의 관계 분석 * 분석을 통해 나온 결과를 근거로 맞춤 대책안 제시 |  |
| 주요기능 | 범죄에 영향을 끼치는 요인을 분석하여 범죄 가능성이 높은 지역과 시간대를 선정하여 순위별로 CCTV나 가로등 설치를 권하고 범죄를 예방할 수 있는 대책을 세우기 위한 구체적 통계자료로 이용할 수 있는 데이터분석 |  |
| 시스템 구성도 |  |  |
| 구축기술 | **1.**Web  - Geometry  - Openlayer  **2.**Web Crawling  - BeautifulSoup  - Selenium  **3.**Parallel Distributed Processing (HDFS)  - ETL(Hadoop) for RDBMS  - Real-Time Processing(Spark) for Web Crawling  [Analysis Technique]  **4.**Classfication and Clustering  **5.**Correlation Analysis |  |
| 소프트웨어 | 1. eclipse  2. Oracle  3. Spark  4. SQOOP  5. HIVE  6. Rstudio  7. ANACONDA : Jupyter  8. QGIS  9. Postgresql |  |