**1페이지 - 인사 및 소개 (1분)**

안녕하세요, 오늘 'DAB4 4팀 Do4mine'에서 준비한 '테러 피해 근절을 위한 글로벌 테러 데이터 분석: 데이터와의 전쟁'이라는 주제로 발표를 시작하겠습니다.

팀장 최지영, 팀원으로는 김동현, 신은서, 발표자 표룡희입니다. 오늘 발표는 테러라는 전 세계적 문제를 데이터 분석의 관점에서 어떻게 접근할 수 있는지, 그리고 이를 통해 어떤 인사이트를 도출할 수 있는지에 대해 다루겠습니다.

**2페이지**

개요입니다. 첫 번째로 프로젝트 소개, 두 번째로 가설 검증. 마지막으로 프로젝트 결론 순으로 말씀드릴 예정입니다.

먼저 프로젝트 소개입니다.

**3,4페이지 - 프로젝트 소개 (2분)**

테러는 아직까지도 전 세계적으로 지속되는 심각한 문제입니다. 9/11 테러 이후 20여 년이 지났음에도 불구하고, ISIS, 알카에다, 탈레반과 같은 테러 조직들은 여전히 전 세계 곳곳에서 무고한 시민들의 생명을 위협하고 있습니다. 특히 내전 지역과 정치적 불안정성이 높은 국가들에서는 테러가 일상화되어 사회적 안정과 경제적 발전을 저해하는 주요 요인으로 작용하고 있습니다.

그렇기에, 저희는 글로벌 테러리즘 데이터베이스(GTD)라는 세계적으로 권위 있는 테러 데이터를 활용하였습니다. GTD는 1970년부터 2021년 6월까지, 약 20만 건의 데이터를 포함하고 있으며, 내용으로는 공격 방식, 목표, 대상, 피해 규모 등 약 135개의 정보를 포함하고 있습니다.

GTD를 이용하여, 테러의 패턴 분석, 테러의 위험성 강조, 민간인 보호를 위한 정책적 시사점을 도출하는 목적을 가지고 프로젝트를 수행하였습니다.

**5페이지 - 데이터 개요 및 전처리 방법 (3분)**

GTD 데이터를 전처리하는 과정에서 우리는 여러 단계의 데이터 클리닝 작업을 수행했습니다. 먼저, df.drop 함수를 사용하여 분석에 필요하지 않은 컬럼들을 제거했습니다.

이 코드를 통해 우리는 다음과 같은 데이터 전처리 작업을 수행했습니다:

1. 미국에 한정한 칼럼의 경우 제외 (글로벌 데이터 분석을 위함)
2. compclaim: 테러 단체 간의 문제이므로 분석 대상의 데이터가 아니므로 제거
3. 데이터와 txt 칼럼 둘 다 있을 경우 데이터 제거
4. divert/kidhijcountry: 납치하여 우회한 국가 → 유의미하다고 판단되지 않아 제거
5. 데이터에 대한 설명 칼럼 및 출처 칼럼 제거

이러한 전처리 과정을 통해 분석에 필요한 핵심 데이터만 남기고, 불필요한 정보는 제거하여 데이터 분석의 효율성을 높였습니다.

**6,7페이지**

이렇게 전처리한 데이터를 토대로, 테러의 양상/데이터 및 원인 분석 후, 가설 검증을 진행하였습니다.

위 사진은 1970년부터 2021년까지 전 세계에서 발생한 테러의 양상으로, 연도별로 지역적 차이가 있다는 것을 확인할 수 있습니다.

**8페이지**

또한, 1970년도부터 테러 발생 건수가 점진적으로 증가했으며, 2010년대 중반에 정점을 찍고 이후, 감소세를 보이고 있습니다.

2015년 이후, 국제 사회 공조 및 IS 약화 등으로 테러 발생 건수는 점차 감소하는 추세를 보이고 있습니다.

이를 단순한 감소세로 해석하기보다는 국제 정세 변화와의 연동성을 분석하는 것이 중요하다 판단하였습니다.

**9페이지**

이를 토대로 6가지의 가설을 세웠으며, 이를 검증해보았습니다.

**10,11페이지 - 가설 검증 1: 내전은 테러 발생에 유의미한 영향을 준다 (2분)**

첫 번째 가설은 "내전은 테러 발생에 유의미한 영향을 준다"입니다. 이 가설을 검증하기 위해 우리는 내전 발생 지역을 Syria, Iraq, Afghanistan, Yemen, Libya, Somalia로 설정하고 분석을 진행했습니다.

카이제곱 통계량과 p-value 분석 결과, p-value가 0.05보다 크기 때문에 내전 여부는 테러 발생 여부에 유의미한 영향을 주지 않는 것으로 나타났습니다. 그러나 추가 분석을 통해, 연도(iyear), 발생 국가(country\_txt), 자살 테러 여부(suicide),

발생 세부 지역(region\_txt), 테러에 사용한 무기 종류(weaptype1\_txt) 등의 변수가 테러 발생에 유의미한 영향을 주는 것으로 확인되었습니다.

결론적으로, 내전 자체보다는 다른 요소들이 테러 발생에 더 중요한 영향을 미치는 것으로 분석되었습니다. 특히 내전이 없더라도 테러는 발생 가능하며, 내전 지역에서의 테러 집계에 어려움이 있어 통계적 외곡 발생 가능성이 존재합니다. 또한 데이터베이스의 낮은 범위 때문에 부족할 가능성도 높아졌습니다.

**12페이지**

**가설 검증 2: 내전 지역에서의 테러는 무장 공격과 폭발/폭격 유형이 주를 것이다 (2분)**

두 번째 가설은 "내전 지역에서의 테러는 무장 공격과 폭발/폭격 유형이 주를 것이다"입니다.

해당 그래프는 각 연도마다 공격 유형의 비율을 나타낸 것입니다.

해당 그래프에서는 1970년대까지는 암살이 주를 이뤘던 반면, 1980년대 이후에는 폭발/폭격 유형이 주를 이루는 것을 알 수 있습니다.

이를 확인하여, 카이제곱 검정으로 추론 통계를 한 결과, p-value가 0.05보다 작아 통계적으로 유의미함이 확인되었습니다. 이를 머신러닝까지 사용하여, 예측하였습니다.

**13페이지**

머신러닝에서는 사망자수 (nkill), 부상자 수(nwound)의 결측치를 0으로 변환 후, 합하여 사상자 수(casualities)라는 새 컬럼을 생성하였습니다

머신러닝의 여러 모델을 사용하였으며, 특히 XGBoost 모델을 사용한 분석에서 Bombing/Explosion의 성능이 가장 높게 나타났으며, Accuracy: 0.84, F1 Score: 0.50의 결과를 보였습니다.

**14페이지**

SHAP 분석 결과에서도 무기 유형(weaptype1\_txt)이 공격 유형을 예측하는 데에 가장 큰 영향을 주는 변수 중 하나임을 알 수 있었습니다. 특히 특정 무기 유형이 있을 경우, 특정 공격 유형으로 예측될 가능성이 높아짐을 보여주었습니다(예: 폭탄류는 폭발 테러로 예측될 가능성이 높아짐).

이를 통해 내전 지역에서의 테러는 무장 공격과 폭발/폭격 유형이 주를 이룬다는 가설이 타당함을 확인했습니다.

**15페이지 - 가설 검증 3: 단체(조직)은 복합공격을, 개인은 단일공격을 중심으로 진행할 것이다 (2분)**

세 번째 가설은 "단체(조직)은 복합공격을, 개인은 단일공격을 중심으로 진행할 것이다"입니다.

분석을 위해 우리는 nperps(가해자수)를 기준으로 다음과 같은 전처리를 수행했습니다:

* nperps가 0, -99, -9(=Unknown) 또는 결측값일 때
* nperpcap(체포인원수) ≥ 10이면 nperpcap값으로 대체
* nperpcap 또한 결측값이라면 행 제거

이러한 방식으로 개인(0), 조직(1)으로 분류하고, 단일공격(0), 복합공격(1)으로 변환했습니다.

**16페이지**

분석 결과, p-value가 0.05보다 작아 통계적으로 유의미함이 확인되었습니다. 조직의 경우 개인에 비해 복합공격을 수행하는 비율이 높고, 전체 비중은 적지만 복합공격 또한 많은 것으로 나타났습니다.

하지만 Cramér's V 값은 0.126으로, 두 변수 간에는 약한 수준의 연관성이 나타났습니다. 이는 복합공격(Multiple Attack) 자체가 매우 드물기 때문에 유의미한 관계는 있지만 효과 크기는 제한적이라는 것을 의미합니다. 따라서 가설이 맞기는 하지만, 그 영향력은 약한 것으로 결론지었습니다.

**17페이지**

테러에서 단일 공격의 비중이 높은 이유를 3가지로 구분할 수 있는데,

실행 비용과 조직력, 실패 확률의 감소, 목표 및 메시지 존재입니다.

**단일 공격은 준비 및 실행이 한 번만에 이루어집니다.**

**하지만, 복합 공격은 여러 공격을 동시에 조율이 가능합니다.**

**그리고 복합 공격은 중간 단계에서 실패하거 발각될 가능성이 높을 것이라 추정됩니다.**

**또한, 특정 물리적 피해보다 실행 자체가 목적일 수도 있기 때문에, 단일 공격으로도 충분히 목적을 달성할 수 있을 것입니다.**

**18페이지 - 가설 검증 4: 전문 테러 조직의 주요 타깃은 정부/군대/경찰일 것이다 (2분)**

네 번째 가설은 "전문 테러 조직의 주요 타깃은 정부/군대/경찰일 것이다"입니다.

분석 결과, 테러가 5회 미만 발생한 표적은 통계적으로 의미가 없어 제외했습니다. p-value가 0.05보다 작아 통계적으로 유의미하지만, 기타 대상의 공격이 더 많으므로 가설 기각되었습니다.

실제로 테러 조직의 표적 유형 분포를 보면, Private Citizens & Property(민간인과 재산)이 24.6%, Military(군대)가 20.8%, Police(경찰)이 14.6%, Business(기업)이 10.3%, Government(일반 정부)이 9.2% 등의 순으로 나타났습니다.

그리고, 정부/군대/경찰을 그룹으로 묶고, 이외의 그룹을 민간 그룹이라고 가정했을 때, 민간 그룹이 약 55%를 차지하는 것을 알 수 있습니다.

**19페이지**

**주요 타겟이 정부 측이 아닐 가능성이 높은 이유도 3가지로 나누어볼 수 있습니다.**

현실적 제약, 민간인 대상의 효과, 테러 조직 내 전략 변화입니다.

* 현실적으로 정부/군대/경찰 그룹은 높은 보안 수준을 유지하고 있을 것입니다.
* 또한, 무장 상태일 가능성이 높습니다.
* 국가적 차원에서의 보복 위험도 큽니다

심리적 공포 확산 및 사회 혼란 조성이 목적이라면, 민간인과 기타를 대상으로 하는 것이 더 효과적일 것입니다.

과거에는 정치적인 목적으로 정부 기관을 타깃으로 노렸지만, 현재는 대중의 노출/경제의 타격/종교적 증오심 등의 목적으로 “soft target” 전략을 선호할 수 있습니다.

**20페이지 - 가설 검증 5: 경제적으로 불안정한 국가일수록 정부 타깃인 테러가 많을 것이다 (2분)**

다섯 번째 가설은 "경제적으로 불안정한 국가일수록 정부 타깃인 테러가 많을 것이다"입니다.

회귀분석을 통해 GDP와 테러 발생 건수의 관계를 분석한 결과, 음의 기울기로 GDP가 높을수록 테러 횟수가 줄어드는 경향을 확인했습니다. p-value가 0.05보다 작아 통계적으로 유의미하지만, 상관계수 r이 -0.1159로 약한 음의 상관관계를 보이며, 결정계수 r²은 0.012로 GDP는 1.2%만 설명 가능함을 보여줍니다.

**21페이지**

따라서 경제적 여건이 테러 발생에 영향을 주는 것은 맞지만, 그 영향이 매우 미미해서 정책적, 사회적 해석에는 거의 의미가 없음을 알 수 있습니다. GDP만으로 테러를 설명하기 어려운 이유로는 GDP는 국가의 경제적 총생산 규모를 나타내는 지표일 뿐, 국민 개개인의 삶의 질, 빈곤 불평, 사회적 억압 등은 반영하지 못하기 때문입니다.

**22페이지 - 가설 검증 6: 정부 기관은 보안이 철저하여 타 그룹보다 테러의 성공률이 낮을 것이다 (2분)**

여섯 번째 가설은 "정부 기관은 보안이 철저하여 타 그룹보다 테러의 성공률이 낮을 것이다"입니다.

분석 결과, 정부 관련 그룹(정부 기관/경찰/군대)의 테러 성공률은 약 87.6%, 그 외의 그룹은 약 92.2%로 확인되었습니다. 또한 p-value가 0.05보다 작아 통계적으로 유의미함이 확인되었습니다.

모델 비교 분석에서는 CatBoost 모델이 가장 정확도가 높았으며, Accuracy: 0.9188, F1 Score: 0.9551의 결과를 보였습니다.

**23페이지**

Military/Police가 최악하고, 그 중 특히 검문소 등이 최약한 것을 알 수 있었습니다. 또한 상위 15개 모든 납치 유형에서 성공률이 높으며, 경찰/군대 조직에서 최약한 것 또한 알 수 있었습니다.

이러한 분석을 통해 정부 관련 기관의 보안이 철저하여 타 그룹보다 테러의 성공률이 낮을 것이라는 가설이 타당함을 확인했습니다.

**24, 25페이지**

**프로젝트 결론 및 제언 (3분)**

프로젝트 요약입니다. 이번 연구를 통해 우리는 테러 발생과 관련된 여러 요인들을 분석하고, 중요한 인사이트를 얻었습니다.

총 3가지로 나누어 설명드리겠습니다.

첫 번째로 가설 1, 4번이 해당되는 기각된 가설입니다.

* 내전은 테러 발생에 유의미한 영향을 주지 않으며, 정부/군대/경찰을 주요 타깃으로 보는 시각도 통계적으로 기각되었습니다.
* 이를 통해, 국가 내부의 혼란이나 체재 반대 성향만으로는 테러 발생을 설명할 수 없으며, 정치적 목적보다 기회주의적 혹은 상징적 동기가 더 강할 수 있음을 나타냅니다.

두 번째로 가설 3, 5번에 해당되는 채택되었으나 상관관계가 부족한 가설입니다.

* 조직은 복합 공격을, 개인은 단일 공격을 선호하며, 경제적 불안정 국가에서 정부 대상 테러가 많다는 가설은 채택되었으나 상관관계가 약했습니다.
* 이를 통해, 공격 주체와 방식 간 선호 경향은 존재하나, 예외가 많으며, 경제적 불안정성 역시 테러를 유발하는 핵심 요인보다는 보조적 배경 변수에 가까움을 알 수 있었습니다.

세 번째로, 가설 2, 6번에 해당되는 가설 채택 후, 머신러닝 학습입니다.

* 내전 지역에서는 폭발/무장 공격이 주로 발생하고, 정부 기관은 보안으로 인해 테러 성공률이 낮다는 가설은 머신러닝 학습에 반영되었습니다.
* 이를 통해, 공격 유형과 타깃의 조합에서 일정된 패턴이 포착되며, 이는 모델링을 통해 위험 예측과 정책 설계에 실질적으로 활용이 가능함을 알 수 있었습니다.

**26페이지**

**이러한 납치 혹은 테러를 줄이기 위하여, 취약성에 대한 프로세스를 구축해야합니다.**

**그렇기에, 시사점 및 제언을 말씀드립니다.**

2014년 런던에서 처음 시작한 Project Servator는 여러 국가에서도 사용하고 있는 정책입니다.

경찰들이 계속 해당 지역을 순찰하며, 해당 지역의 거주자와 기업의 지원, CCTV 등의 지원을 받아, 의심스러운 활동을 감지하도록 참여하는 전술입니다.

항상 경찰들이 지역을 순찰하고 있기에, 상황 발생 시, 바로 전달할 수 있어, 해당 상황을 빠르게 처리할 수 있습니다.

이를 타 지역, 혹은 타 국가에도 구축한다면 시민들도 테러 혹은 납치 등의 사건을 미연에 방지할 수 있다 생각합니다.

**27페이지**

그리고 가설 2를 토대로, 내전 지역에서는 폭발/폭격, 무장 공격 중심이 되는 테러가 많기 때문에, 무기 밀반입 차단과 공공장소의 보안 강화가 필수적입니다.

특히, 공항, 지하철, 군사시설 등과 같은 주요 시설에 대해서는 무기 및 폭발물 탐지 장비의 고도화된 고위험 무기 유형에 대한 집중 감시 정책을 마련해야합니다.

예멘의 경우, 2010년, 알카에다의 위협에 대응하여, 공항 보안을 대폭 강화한 사례가 있습니다.

최신식 스크린 장비와 다양한 폭발물을 탐지할 수 있는 장비를 설치하고, 승객들에 대한 검색을 철저히 수행한 사례가 있습니다.

**28,29페이지**

**향후 연구 과제 제시**

1. 도시 설계 및 건축물 특성 데이터, CCTV 및 경비 데이터, 인구 밀도 및 유동 데이터들과 GTD의 시계열 데이터를 활용하여, 도시 공간의 특성과 테러 발생 패턴에 대한 공간 분석 연구를 하여**, "방어적 도시 설계 전략이 테러 억제에 미치는 효과 측정”**을 해볼 수 있을 것입니다.
2. CERT/CSIRT의 사이버 공격 데이터, 다크웹 모니터링 데이터, 국가별 사이버 보안 준비도 지수 등을 활용하여, 사이버 테러 물리적 테러 사이의 선후 관계 및 상관성을 분석하여 **"하이브리드 테러 위험에 대한 보안 프레임워크 개발"**을 해볼 수 있을 것입니다.

이러한 연구를 통해 우리는 데이터 기반의 테러 예방 및 대응 전략을 개발하고, 안전한 사회 구축에 기여할 수 있을 것입니다. 특히 기존의 전통적인 보안 접근법을 넘어, 빅데이터와 인공지능을 활용한 선제적 위험 탐지 및 예방 시스템 구축이 필요합니다.

**30페이지 - 마무리 (1분)**

발표를 마치며, 테러는 단순한 안보 문제를 넘어 사회적, 경제적, 정치적 요인이 복합적으로 작용하는 현상임을 강조하고 싶습니다. 데이터 분석을 통해 우리는 테러의 패턴, 대상, 방법 등에 대한 객관적 인사이트를 얻을 수 있었고, 이는 효과적인 대응 전략 수립에 중요한 기초가 될 것입니다. 감사합니다.