(p. 1) 안녕하세요, “사고위험지수 개발을 통한 강서구의 보행 안정성 제고 방안” 발표를 시작하겠습니다.

(p. 2) 목차는 시간관계 상 생략하겠습니다.

(p. 3) 저희 팀은 강서구에서 지난 5년간 발생했던 교통사고 발생건수와 사망사고를 확인해보았습니다. 먼저 왼쪽 그래프를 보시면, 2019년부터 2021년까지 강서구는 서울시에서 다섯 번째로 많은 교통사고 발생을 기록하고 있었습니다.

또한 오른쪽 그래프를 보시면 알 수 있듯, 강서구에서 교통사고로 발생한 사망자수는 2019년과 20년에 각각 서울시 내 1위를 기록하였고, 21년도에는 5위를 기록하였습니다. 이처럼 강서구에서는 교통사고가 자주 발생하면서도, 보행자에게 심각한 수준의 사고도 많이 일어남을 확인하였습니다.

(p. 3, 애니메이션) 그래서 저희 팀은 강서구의 교통사고 발생량을 줄임으로써 보행 안전성을 확보하고 싶었습니다. 이를 위한 방안을 고민했을 때, 보행자에게 현재 위치에 대한 교통사고 위험수준을 전달하여 경각심을 부여하면 효과적이지 않을까 생각했습니다.

(p. 4) 기존에도 교통안전과 관련한 지표가 존재했습니다. 도로교통공단에서 제안한 교통안전지수가 바로 그것인데요. 안타깝게도, 이 지표는 지자체 전체를 평가한 하나의 점수와 같았습니다. 즉, 세부적인 위치에 대한 지표는 아니었는데요. 저희는 좀 더 세부적인 지역에 대해 사고 위험을 나타내면 좋을 것이라고 생각했습니다.

이를 위해서는 세부적인 지역을 명확하게 정의할 필요가 있었습니다. 저희는 횡단보도를 세부지역으로 정의하고, 이를 기준으로 한 지수를 만들고자 했습니다. 그 이유는, 바로 보행자 교통사고의 무려 50%가 횡단 중에 발생하였기 때문입니다. 궁극적으로, 저희 팀은 강서구 내의 횡단보도별로 교통사고 위험수준을 나타내는 지표를 만들어서 보행안정성을 제고하는 것을 분석 목표로 설정하였습니다.

(p. 5) 저희의 분석을 위해서는, 각각의 횡단보도 ID를 기준으로 타겟변수가 될 교통사고 발생건수와 독립변수로 사용할 시설물 개수 등을 집계한 데이터프레임이 필요했습니다. 하지만, 기존의 데이터 중 그러한 데이터는 없었습니다. 그래서 저희는 횡단보도를 기준으로 한 데이터프레임을 직접 만들었습니다.

이를 위해 먼저, 총 13개의 강서구 공공데이터를 수집하였습니다. 그리고 횡단보도 좌표와 시설물좌표를 기준으로 개별 시설물 등의 거리를 계산하여 각 횡단보도 근처의 시설물 개수를 데이터프레임으로 만들었습니다. 좌표와 좌표 사이의 거리를 계산하는 데에는 하버사인 거리를 사용하였습니다.

(p. 6) 하버사인 거리란, 지구의 곡률을 고려하여 거리를 계산한 방법입니다. 더 정확한 거리 계산을 위해서 이 방법을 고려하였습니다. 오른쪽 그림을 보시면, 횡단보도의 위경도 좌표를 기준으로 반경 100m를 검은 원으로 표시하였습니다. 그 안에 빨간 점은 교차로 데이터의 좌표를 표시한 것인데요. 이처럼 검은 원 안에 포함되는 시설물을 집계해서 데이터프레임으로 만든 것입니다.

(p. 7) 다음은 저희가 분석에 사용한 설명변수의 대표적인 예시입니다. 뉴스에 따르면, 불법주정차가 교통사고 위험을 높이는 한 요소가 될 수 있어서 고려하였습니다. 교차로 역시 마찬가지입니다.

그리고 유동인구를 고려하고 싶었는데, 공공데이터로는 세부적인 위치에 대한 유동인구를 파악할 수 없었습니다. 그래서 저희는 지하철역과 버스정류장에서 승하차하는 인구로 유동인구를 대체하고자 하였습니다. 정류장 거리를 기준으로 가중치를 서로 달리하여 가중치가 반영된 수치를 계산하였습니다. 이외에도 다른 설명변수가 있는데, 횡단보도를 기준으로 몇 m를 기준으로 했는지는 오른쪽 아래 표에 정리된 바와 같습니다.

(p. 8) 최종적으로 강서구 내의 2,442개 횡단보도에 대해 총 13개의 칼럼을 갖는 데이터프레임을 완성하였습니다.

(p. 9) 완성된 데이터프레임을 사용하여 회귀분석을 진행했습니다. 타겟변수는 각 횡단보도의 교통사고발생건수를 사용하였고, 나머지 11개 변수를 독립변수로 하여 어떤 변수가 유의한지 변수선택을 진행했습니다. 그 결과, 스마트횡단보도, 학원 수 가 유의하지 않아서 이 변수를 제외하고 총 9개 변수를 사용하여 분류를 진행했습니다.

(p. 10) 다음으로, 앙상블 기법을 활용해 분류를 진행했습니다. 기존 타겟변수는 교통사건발생건수로 연속형 변수였습니다. 이를 분포를 확인한 후 ‘교통사고 거의 없음’, ‘간혹 발생’, ‘자주 발생’ 총 3개 범주를 갖는 범주형 변수로 변환하였습니다. 그리고 변환한 범주형 변수로 분류 모형을 적용하여, 랜덤포레스트, xgboost, 그래디언트 부스팅 모형을 적용한 결과, 정확도가 가장 높은 xgboost를 최종 모형으로 사용하기로 결정했습니다.

(p. 11) 그리고 xgboost의 변수 중요도를 확인해보았습니다. 주변 학교유무, 교차로 유무, 지하철 승하차 인구 가중값 변수 순으로 변수 중요도가 높게 나타났습니다. 저희는 계산된 변수 중요도를 사고위험지수 산출에 가중치로 활용하기로 하였습니다. 그 이유는, 변수 중요도가 나타내는 의미가 바로 각 변수가 분류에 얼마나 영향을 미치는지 나타내는 지표이기 때문이었습니다. 그래서 이 변수 중요도를 가중치로 산출식을 개발하였습니다.

(p. 12) 계산된 사고위험지수 산출식은 다음 변수 중요도의 값들을 사용하였습니다. 이 지수를 선형적으로 사용하여 0에서 100 사이의 값을 갖는 ‘사고위험지수’를 생성하였습니다.

(p. 13) 저희가 생성한 ‘사고위험지수’가 타당한 수치인지 확인하고자 분산분석을 실시하였습니다. 각 그룹이 유의한 차이를 나타냄을 확인할 수 있었습니다. 그리고 왼쪽 위 그래프를 보시면 density plot을 보실 수 있는데, 각 선마다 봉우리의 위치가 크게 차이남을 볼 수 있었고, 또한 박스 플롯을 확인했을 때도 그룹간 비교적 명확한 차이를 보였습니다. 따라서 해당 지수가 사고위험을 충분히 나타낼 수 있는 값이라고 생각되었습니다.

(p. 14) 저희가 산출한 사고위험지수를 대시보드로 나타내 보았습니다. 왼쪽 결과를 보시면, 사고위험지수가 83으로 매우 높은 교통사고 위험도를 나타내고 있는 지역입니다. 실제로 해당 지역은 교통사고 ‘자주 발생’에 해당하는 지역이었고, 이 지역의 특징으로는 주변에 학교가 있으며, 유동인구가 많은 편이고, 불법주정차 단속건수가 많은 편임을 알 수 있었습니다.

반대로, 오른쪽 그림을 보시면, 사고위험지수가 24.79로 나타나는 지역입니다. 해당 지역은 실제 사고발생빈도가 ‘거의 없음’에 해당하는 지역이었고, 이 지역의 특징으로는 주변에 학교가 없었으며, 유동인구가 적었고, 불법주정차 단속건수가 0건인 지역이었습니다.

이처럼 저희가 개발한 사고위험지수를 강서구 내의 세부적인 지역의 교통사고 위험수준을 판별하는 기준으로 활용할 수 있을 것으로 보입니다.

(p. 15) 마지막으로, 저희가 개발한 사고위험지수를 활용하기 위한 방안을 고려해 보았습니다. 횡단보도 증축 및 이전 시 사고위험도를 고려하여 횡단보도 위치 추천에 활용하거나, 위험도가 높은 횡단보도를 우선적으로 보수하는 방안을 고려해 보았습니다. 그리고 해당 지수를 도로교통안전 정책의 자료로 활용할 수 있을 것으로 생각하였습니다.

또한, 저희가 가장 중요하게 생각했던 방안으로 사고위험지수를 지도 서비스 중 하나로 추가하여 시민에게 안전 정보를 제공하는 방안과 보행신호 음성안내 보조장치로 사고위험지수를 통한 위험 안내도 함께 고려해 보았습니다.

(p. 16, 끝) 저희의 사고위험지수를 활용해서 시민들이 보다 안전한 보행환경을 누릴 수 있을 것으로 기대하며 발표 마치겠습니다. 감사합니다.