

범죄 데이터 기반 범죄 예측 및 순찰 알고리즘 개발

Crime Data-Based Crime Prediction and Patrol Algorithms

류형주 팀장

2016112650

김동연

2016112622

김영서

2017112596



Index



주제 소개



주제 소개

- 목적

기존의 순찰경로

주민 요청이나 경찰관의 감으로
순찰경로 구성

새로운 순찰경로

다양한 데이터를 활용해 범죄 위험 지역을 예측하고
해당 결과를 바탕으로 최적의 순찰경로를 구성함

‘범죄 발생 시 **초동조치** 시간이 빠름’

주제 소개

• 데이터 분석 기반 범죄 예측

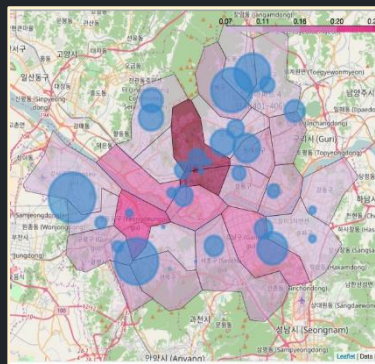
5대 범죄 발생 현황
(단위: 건, 매년 1~5월 누계 기준)

	2008	2009	2010	2011	2012년
폭력	9만 5563	10만 1656	9만 2494	11만 6524	11만 6857
절도	8만 565	9만 5981	9만 9026	12만 6121	10만 1355
강간	5001	5323	6767	7320	6353
강도	1861	2793	2004	1803	1143
살인	396	489	491	541	422

자료: 경찰청

범죄 데이터

- 5대 범죄 유형, 발생 건수
- 발생 시간대, 날짜 특성 고려



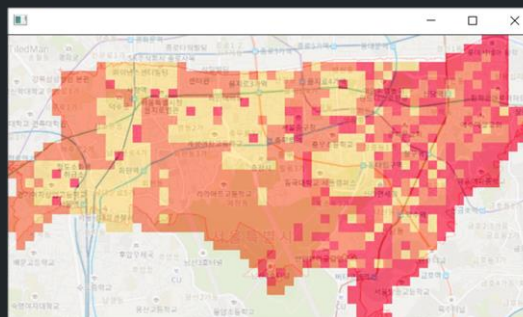
위치 데이터

- 범죄 취약 지역 분석 (서울특별시 중구 한정)
- 인근 경찰서까지 거리, CCTV개수



상황 데이터

- 범죄 발생 당시 날씨
- 지역의 유동인구 및 거주인구



지역별 순찰 중요도 도출

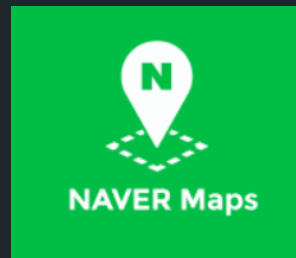


주제 소개

• 범죄 예측 기반 순찰 알고리즘 개발

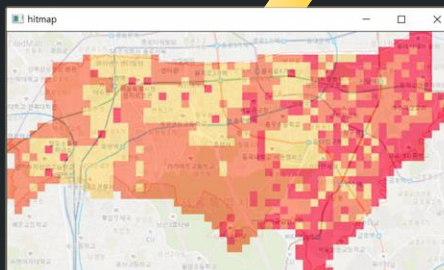


기존 순찰 경로



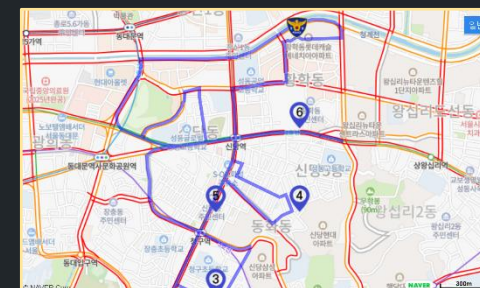
관할 구역 교통 데이터 활용

- 시간대별 교통량, 신호주기 이용
- NAVER Maps API 활용



지역별 순찰 중요도

- 데이터 기반 범죄 발생 위험 지역 예측
- 위험 지역을 고려하여 순찰 가능하도록 알고리즘 개발



데이터 분석 기반 순찰 경로

- 범죄 발생 위험지역 우선 고려
- 사건 발생 원점까지 거리 단축



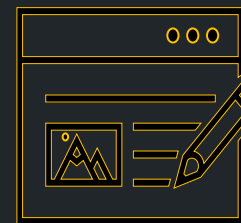
Index



Step. 1
범죄 예측 지도 생성



Step. 2
순찰 경로 알고리즘 개발



프로젝트 결과



Step.1 : 범죄 예측 지도 생성

- 데이터 수집 – 위치, 상황 데이터



날씨 데이터

시점	일시	기온(℃)	강수량(mm)	풍속(m/s)	습도(%)	일조(%)	적설(cm)	한온(°C)
108	2019-01-01 0:00	-5.5	1	54				0
108	2019-01-01 1:00	-5.9	1.8	56				0
108	2019-01-01 2:00	-6.5	1.2	60				0
108	2019-01-01 3:00	-6.9	2.2	62				1
108	2019-01-01 4:00	-7.2	1.2	57				3
108	2019-01-01 5:00	-7.6	2.1	59				1
108	2019-01-01 6:00	-7.9	1.3	60				4
108	2019-01-01 7:00	-7.7	1.3	58				6
108	2019-01-01 8:00	-7.7	0.9	57	0			8
108	2019-01-01 9:00	-7	0.9	50	0			9
108	2019-01-01 10:00	-4.9	1.6	44	0.6			8
108	2019-01-01 11:00	-3.7	2.5	41	1			5
108	2019-01-01 12:00	-2.8	2.8	47	1			8
108	2019-01-01 13:00	-1.5	2.7	49	1			7
108	2019-01-01 14:00	-1.7	3.5	51	1			8
108	2019-01-01 15:00	-1.5	3	46	0.8			4
108	2019-01-01 16:00	-0.9	3.2	45	1			2
108	2019-01-01 17:00	-1.8	2.2	48	1			2
108	2019-01-01 18:00	-2.7	2.6	45	0.1			1
108	2019-01-01 19:00	-3.2	3.2	45				0
108	2019-01-01 20:00	-3.9	3.5	42				1
108	2019-01-01 21:00	-4.6	2.6	36				0
108	2019-01-01 22:00	-5.4	3.1	43				2



CCTV, 경찰서 데이터

개소 관리번호	주용도	설치대수	Latitude	Longitude
공002	공원 방법	1	37.55567	126.9645
공005	공원 방법	1	37.5527	127.0168
공006	공원 방법	1	37.56017	126.9721
공007	공원 방법	1	37.56017	126.9721
공011	공원 방법	2	37.56686	126.9708
공012	공원 방법	2	37.56686	126.9708
공013	공원 방법	2	37.56428	126.9685
공014	공원 방법	2	37.56778	126.9842
공015	공원 방법	2	37.56814	126.9807
공016	공원 방법	2	37.56814	126.9807
공017	공원 방법	2	37.56441	127.005

지역	관할서	기관명	latitude	longitude
서울청	서울중로경찰서	새종로파출소	126.9763591	37.5689804
서울청	서울중부경찰서	광희지구대	127.0134056	37.5652525
서울청	서울중부경찰서	서울중부경찰서	126.9896139	37.5635144
서울청	서울중부경찰서	신당파출소	127.0163928	37.5650526
서울청	서울중부경찰서	역수지구대	127.012375	37.552175
서울청	서울중부경찰서	율지로3가파출소	126.9828763	37.5661306
서울청	서울중부경찰서	율지구대	127.0038758	37.5665337
서울청	서울중부경찰서	강종파출소	127.0047941	37.55877
서울청	서울중부경찰서	종무파출소	126.9915414	37.5608047
서울청	서울남대문경찰서	남대문파출소	126.9764128	37.559012
서울청	서울남대문경찰서	영동파출소	126.9839591	37.5636434
서울청	서울남대문경찰서	서소문파출소	126.9754723	37.5638243



인구 데이터

구분	단위(2017.6.30)				세대수
	계	남	여		
분령	123,970	61,343	62,627		59,368
소공동	1,979	983	996		1,022
혁현동	5,381	2,947	2,434		3,269
명동	3,007	1,272	1,735		1,453
필동	4,076	2,089	1,987		2,244
창충동	4,956	2,473	2,383		2,969
광희동	5,077	2,684	2,393		2,958
율지로동	1,888	1,059	629		1,194
신당동	7,946	4,015	3,831		4,351

일자	시간	연령대	성별	시	군구	유동인구수
20190601	00	20	남성	서울	강서구	98200
20190601	00	20	남성	서울	금천구	17260
20190601	00	20	남성	서울	서대문구	26880
20190601	00	20	여성	서울	강북구	27920
20190601	00	20	여성	서울	강북구	20340
20190601	00	20	여성	서울	성북구	22010
20190601	00	20	여성	서울	성북구	35460
20190601	00	20	여성	서울	영광구	44580
20190601	00	30	남성	서울	강남구	53780
20190601	00	30	남성	서울	광진구	50790
20190601	00	30	남성	서울	노원구	56830
20190601	00	30	남성	서울	동대문구	27720
20190601	00	30	남성	서울	동작구	54330
20190601	00	30	남성	서울	마포구	56690
20190601	00	30	남성	서울	서초구	55540
20190601	00	30	남성	서울	은평구	55130

- 기상청 제공 원본 데이터 수집
- 2014년 ~ 2019년 1시간 단위 관측 데이터
- 기온, 강수량, 풍속, 습도 등 필요한 데이터 정제

Ranson(2014), Horrocks and Menclova(2011), Elleng and Cohn 1990)

중구 CCTV, 경찰서 위치 데이터 수집

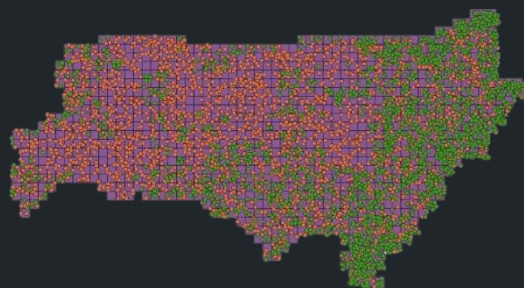
행정동 별 유동인구, 거주인구 데이터 수집

Reginald E. Watts (2012)

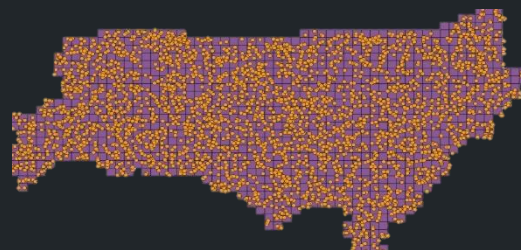


Step.1 : 범죄 예측 지도 생성

- 데이터 수집 – 범죄 데이터



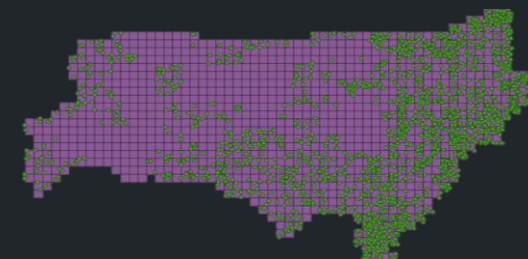
=



랜덤 데이터

- 2014년 ~ 2019년 5대 범죄 통계 기반
- 시간대별 범죄 발생 통계 기반

+



경향성 데이터

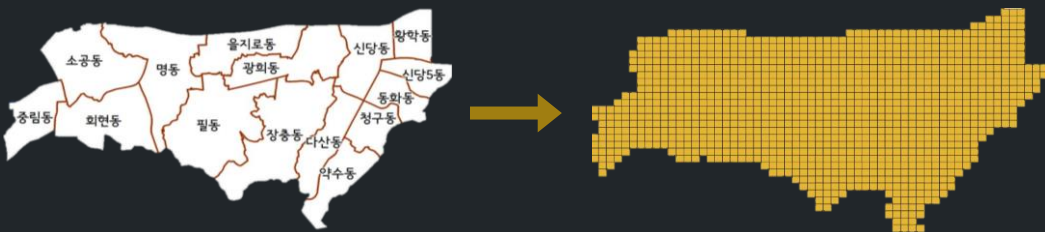
- CCTV ↓ 경찰서 ↓ 유동인구 ↑ 거주인구 ↑ 그리드 선별
- 연도별 2000건의 범죄 사례 추가

실제 데이터 수집에 어려움이 있음(보안상)
2014-2019년 5대 범죄 발생횟수 통계와 시간대
별 범죄 통계를 반영하여 데이터를 생성



Step.1 : 범죄 예측 지도 생성

- 데이터 수집



서울특별시 중구 지도

배열(1107개의 Grid)로 변환

DB

id	type	date_time	latitude	longitude	grid
1	Rape	2014-07-23 08:00:00	37.5578439600	126.9747400000	168
2	Theft	2014-06-17 03:00:00	37.5662467400	127.0062234000	728
3	Theft	2014-05-01 14:00:00	37.5614126200	126.9757099000	164
4	Theft	2014-04-08 05:00:00	37.5581181600	126.9738549000	148
5	Theft	2014-06-22 06:00:00	37.5493602500	127.0064423000	747
6	Theft	2014-11-03 19:00:00	37.5574481900	127.0213194000	1064
7	Theft	2014-09-12 11:00:00	37.5685800700	127.0142751000	915
8	Theft	2014-06-22 04:00:00	37.5626873100	127.0256731000	1098
9	Assault	2014-10-02 22:00:00	37.5695822700	126.9780351000	211
10	Assault	2014-07-31 17:00:00	37.5687341600	126.9769519000	194
11	Theft	2014-07-06 11:00:00	37.5550613500	126.9613399000	3
12	Assault	2014-08-18 00:00:00	37.5665329600	127.0157856000	961
13	Assault	2014-12-18 17:00:00	37.5621660100	126.9682399000	61
14	Theft	2014-12-16 23:00:00	37.5530262500	126.9813612000	267
15	Theft	2014-09-12 13:00:00	37.5526424700	127.0150902000	956
16	Assault	2014-07-06 10:00:00	37.5589623200	127.0063111000	737
17	Assault	2014-09-15 22:00:00	37.5706127600	127.0201754000	1033
18	Assault	2014-08-07 10:00:00	37.5559498700	126.9965221000	551
19	Theft	2014-02-02 09:00:00	37.5686989700	127.0246052000	1086
20	Assault	2014-06-02 14:00:00	37.5586067700	127.0104233000	852
21	Theft	2014-08-23 14:00:00	37.5555085900	127.0003773000	622
22	Theft	2014-03-20 22:00:00	37.5640366100	126.9757695000	180
23	Theft	2014-11-23 22:00:00	37.5492698900	126.9948762000	508
24	Theft	2014-05-05 18:00:00	37.5574290700	126.9876207000	373
25	Assault	2014-07-23 17:00:00	37.5692791500	126.9989818000	584
26	Theft	2014-03-24 16:00:00	37.5680841200	127.0191706000	1018
27	Assault	2014-04-28 00:00:00	37.5630193900	127.0208130000	1041
28	Theft	2014-03-04 03:00:00	37.5600903100	127.0143438000	924



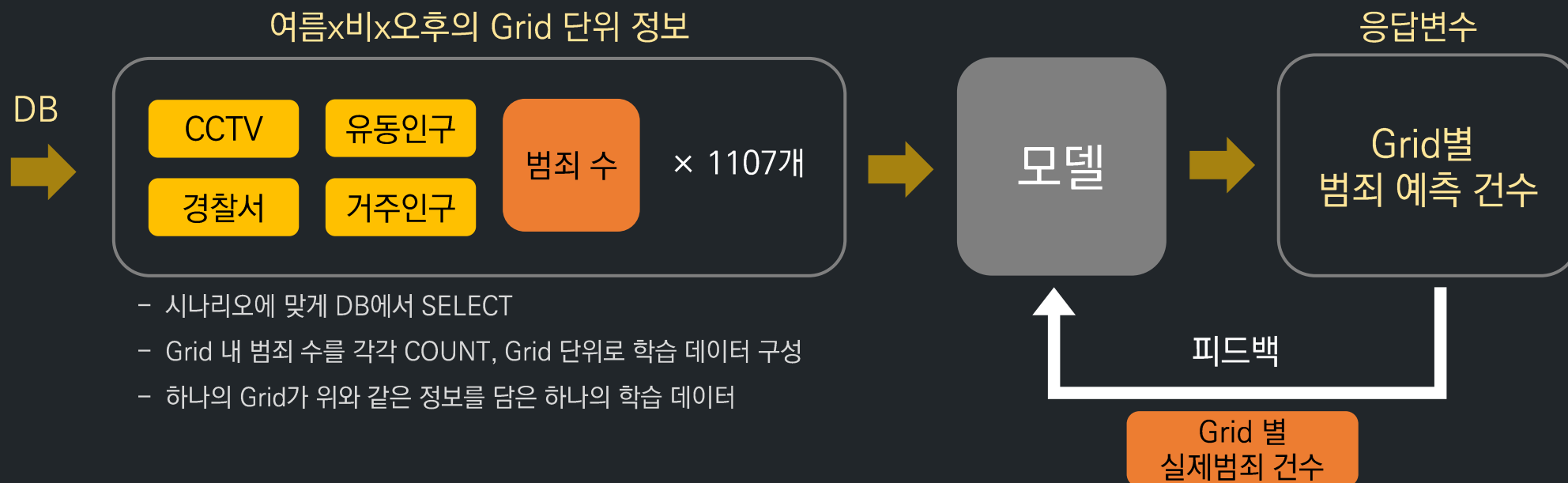
Step.1 : 범죄 예측 지도 생성

- 데이터 분석 - 모델 시나리오 설정

모델 시나리오 설정

여름(계절), 비 오는 날(날씨), 오후(시간)에 어느 지역이
범죄위험이 높은 지 판단하는 모델을 만들자!

* (봄, 여름, 가을, 겨울) × (아침, 오후, 저녁, 심야) × (눈/비, 맑음, 흐림) = 48개 모델 종류





Step.1 : 범죄 예측 지도 생성

- 지도학습 모델 선택

블랙박스 알고리즘

의사결정 나무
(Decision Tree)

랜덤 포레스트
(Decision Tree)

데이터,
매개변수에 민감

서포트 벡터 머신
(SVM)

신경망
(Neural Network)

TABLE 3. Response Plot by Primary Model



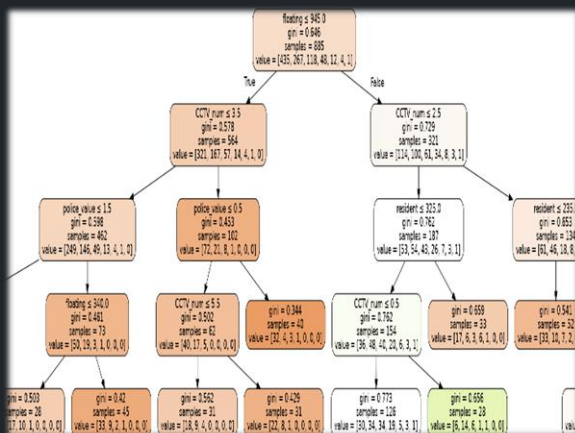
[논문] 머신러닝 기반 범죄발생 위험지역 예측 * 허선영 · 김주영 · 문태현



Step.1 : 범죄 예측 지도 생성

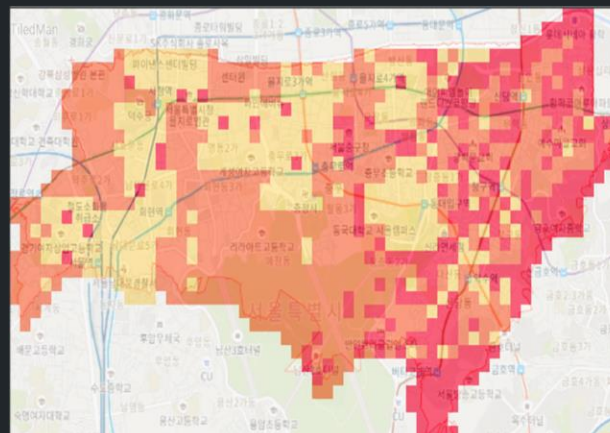
- 데이터 분석 - 분석 결과 시각화

의사결정나무



- DB의 데이터셋을 이용해 학습
- K-Fold 교차검증, 최적 하이퍼 파라미터 설정

그리드별 범죤 예측



- **그리드 별 범죄 예측 결과**

예측 결과 평가

		Predictions											
		1	2	3	4	5	22	23	
Actual class	1	23	20	0	3	0	0	0	
	2	1	40	1	2	4	4	0	
	3	0	3	33	13	2	2	0	
	4	0	6	10	34	5	0	0	
	5	2	0	0	0	41	0	0	
	
	
	
	
	
22	0	0	0	1	0	20	2		
23	0	1	0	0	0	34	1		

- Confusion Matrix를 응용
- 기준 값 이하일 시 모델 재분석

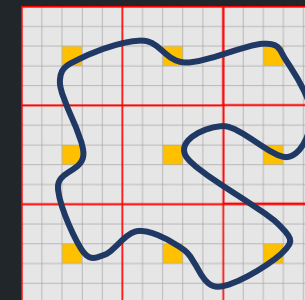
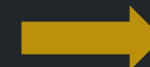
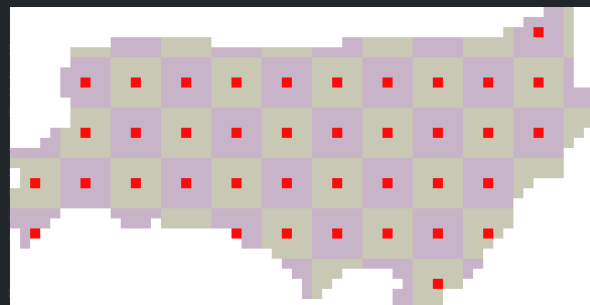
Step.2 : 순찰 경로 알고리즘 개발

알고리즘 설계 - 순찰 지점 추출

TSP
(Traveling Salesman's Problem)

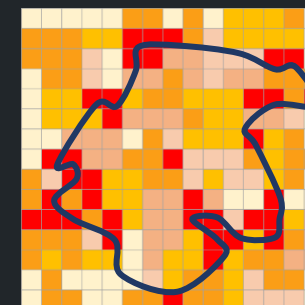
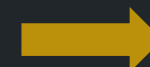
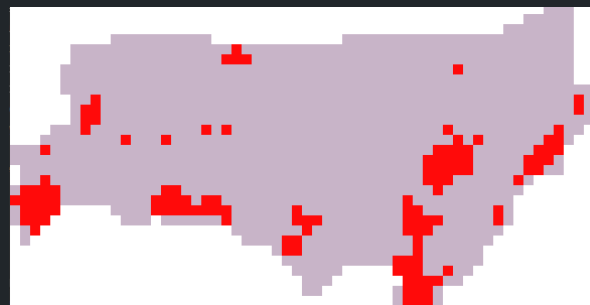
알고리즘 A

- 구역을 나누고 중앙 지점을 순찰해, 전체 지역을 순회하는 것에 집중한 경로
(범죄 예측이 포함되지 않은 대조군)



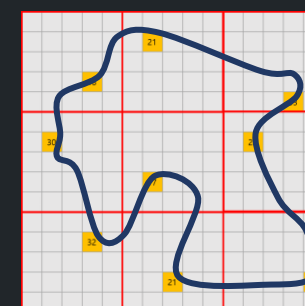
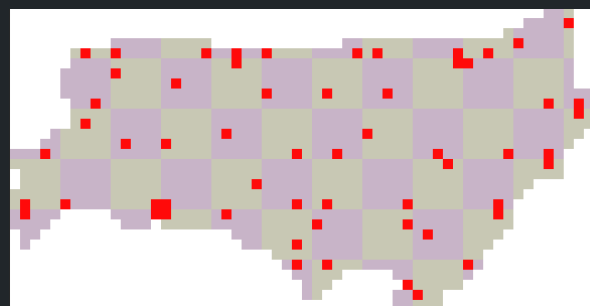
알고리즘 B

- 예측 결과 범죄 고위험 지역을 순찰
(경로가 일부 지역에만 편중될 수 있음)



알고리즘 C

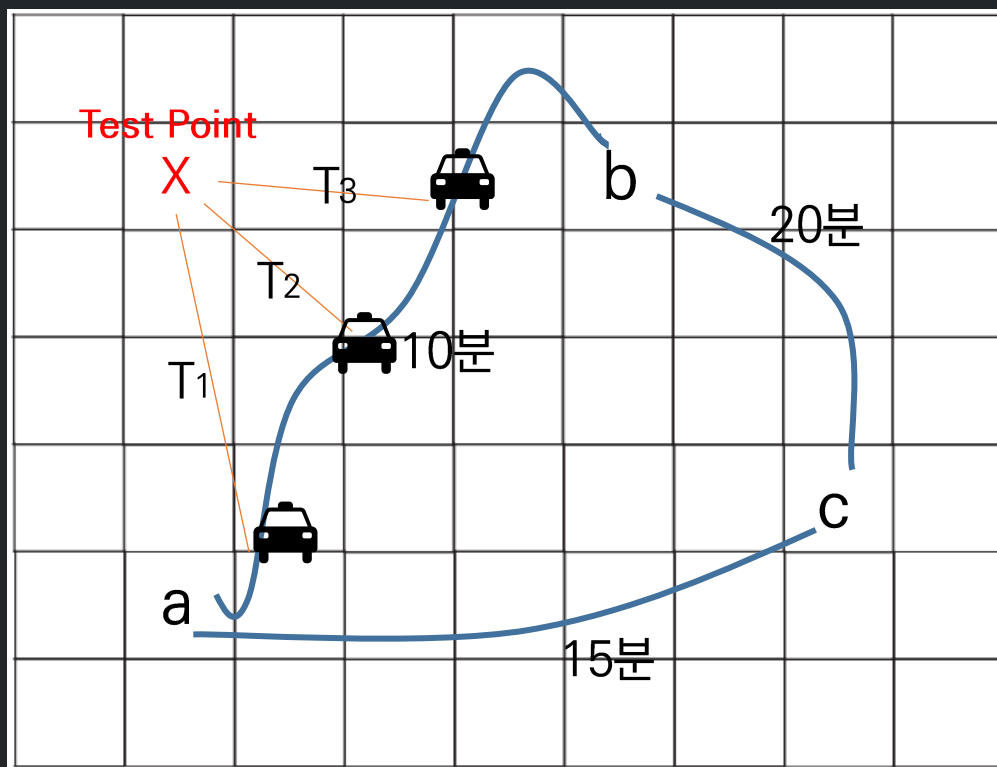
- 구역을 나누고, 구역 내 범죄 위험도가 가장 높은 그리드를 순회하는 경로



Step.2 : 순찰 경로 알고리즘 개발

• 알고리즘 성능 평가 방법

- 사건 발생 지점까지 순찰차의 초동조치 시간을 측정



- 순찰차의 위치
 - 부분경로의 이동시간에 비례하여 순찰차의 위치 결정
 - 범죄 발생 시 a→b 경로에 있을 확률

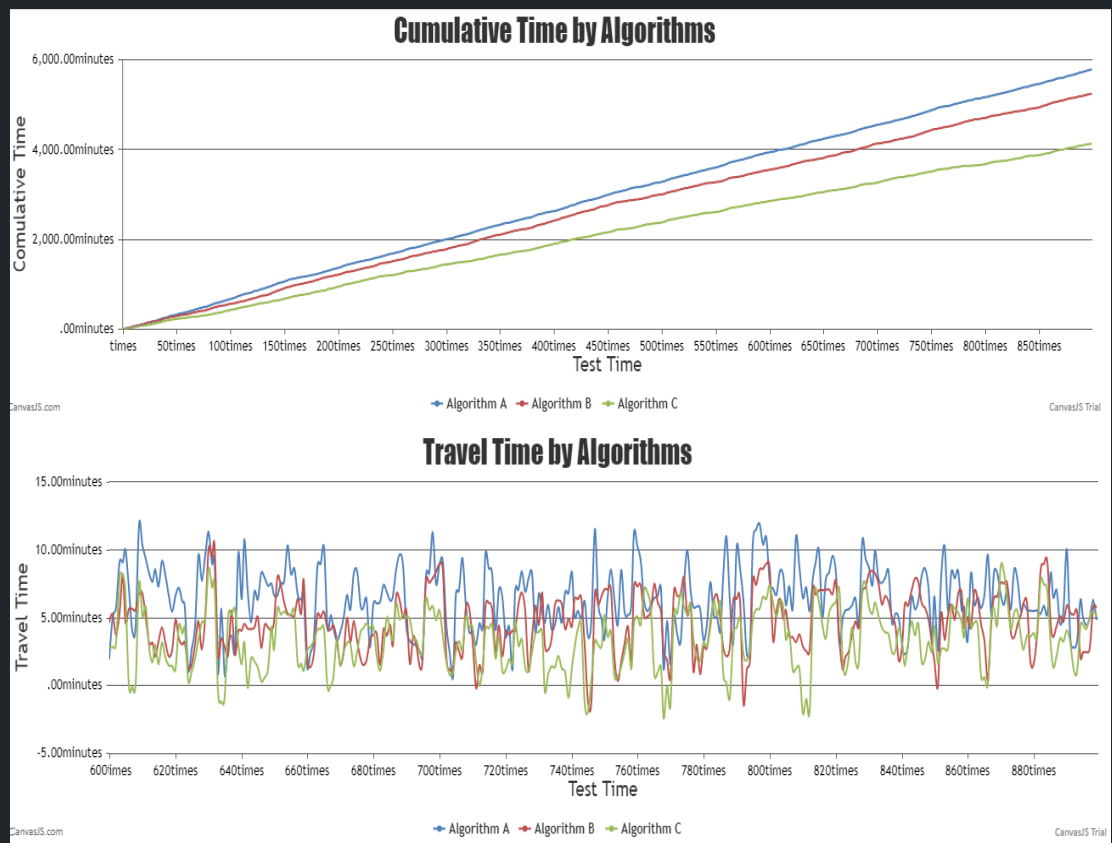
$$\frac{10\text{분}}{10\text{분} + 20\text{분} + 15\text{분}}$$
- 초동조치 시간
 - 경로 내 Test Point를 선택
 - 범죄 발생 지점 까지 걸린 초동조치 시간 누적
- Test Points
 - 실제 데이터와 유사하게 임의 사건 발생
 - 관할 구역당 300개의 Test Point 생성
 - 선택된 부분 경로에서 임의의 3개 지점 추출



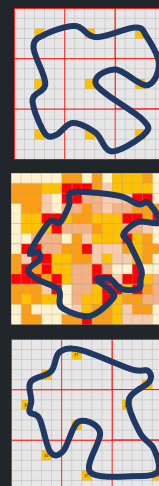
관할 구역별 알고리즘당 900번 Test

Step.2 : 순찰 경로 알고리즘 개발

알고리즘 성능 평가 결과



봄x맑음x저녁 시나리오 Simulation 결과



A 알고리즘

: 범죄 예측이 고려되지 않은 대조군

B 알고리즘

: 고위험 지역만을 순찰하는 대조군

C 알고리즘

: 구역별 고위험 지역을 순찰하는 알고리즘

	1	2	3
A 알고리즘	399	785	1516
B 알고리즘	819	1147	734
C 알고리즘	1482	769	449

알고리즘별 초동조치 누적 순위

프로젝트 결과

범죄 예측 및 순찰 알고리즘

crime2.dothome.co.kr 범죄 예측 및 순찰 알고리즘

War Against Crime

범죄 데이터 기반 범죄 예측 및 순찰 알고리즘

Crime Data Based,
Crime Prediction & Patrol Algorithm

계절

시간

날씨

구역

실행

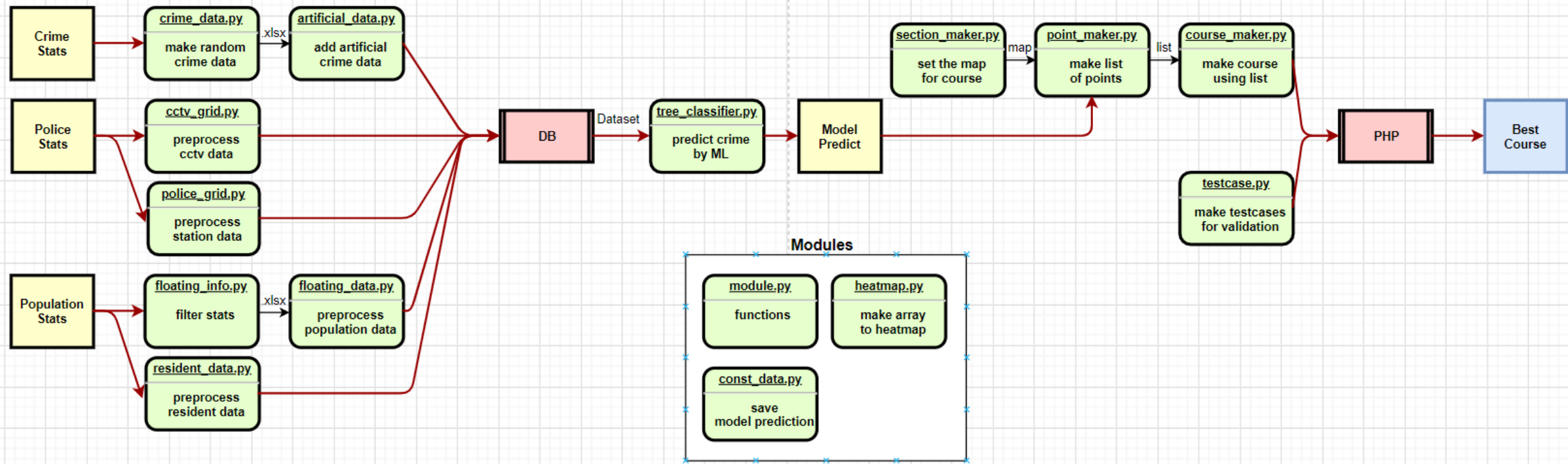
Academic Bases



역할 분담

- 역할 분담

Team ‘범죄와의 전쟁’		
류 형 주	김 동 연	김 영 서
데이터 수집 및 가공		
데이터 생성 & 분석 결과 시각화	DB 설계 & 데이터 분석	데이터 생성 & 데이터 분석
순찰 경로 알고리즘 개발	알고리즘 평가 환경 구축	순찰 경로 알고리즘 개발
알고리즘 검증 테스트 & UI 개발		



An abstract graphic on a dark gray background. It features three concentric dashed circles. The largest circle has four colored dots on its circumference: a blue dot at the top, a light green dot at the top-right, a red dot at the right, and a white dot at the bottom-right. The text 'Q&A' is centered in a yellow, serif font.

Q&A

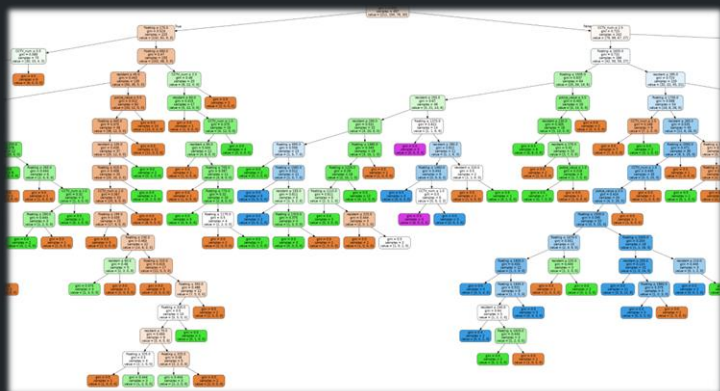
The image features a dark gray background with three concentric dashed circles. Four colored dots are positioned on the outermost circle: a blue dot at the top, a light green dot slightly to the right, a red dot further to the right, and a white dot at the bottom right. The Korean text '추가 자료' is centered in the middle of the image.

추가 자료

데이터 분석

• 의사결정 나무

과적합 의사결정나무



- 가을x심야x맑음 모델에서 실제 과적합이 일어난 모습
- 의사결정 나무는 과적합이 일어나기 쉬움
- K-Fold 교차 검증 사용 및 최적 하이퍼 파라미터 설정을 통해 개선

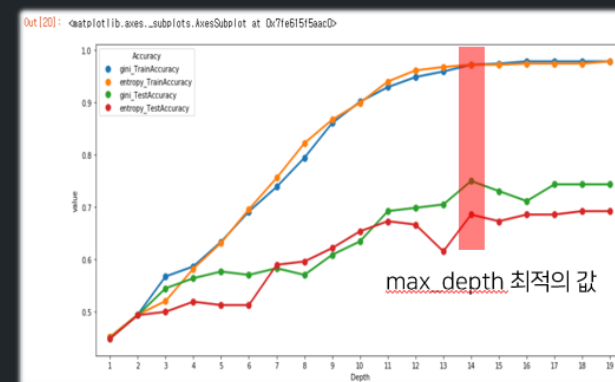
K-Fold 교차검증

Train Data					Test Data
Fold 1	Fold 2	Fold 3	Fold 4	Validation 5	Validation score #1
Fold 1	Fold 2	Fold 3	Validation 4	Fold 5	Validation score #2
Fold 1	Fold 2	Validation 3	Fold 4	Fold 5	Validation score #3
Fold 1	Validation 2	Fold 3	Fold 4	Fold 5	Validation score #4
Validation 1	Fold 2	Fold 3	Fold 4	Fold 5	Validation score #5
Train Data					Test Data

TestSet을 이용한 최종검증

Final Score Average

최적 하이퍼 파라미터 설정



데이터 분석

• 범죄 예측 평가방법

혼동 행렬(Confusion Matrix) 응용

- 값을 정확하게 분류하는 것보다 범죄 경향성을 파악하는 것이 중요
(어느 정도의 범위 안에 들도록 예측하는 것이 목표)

		Predictions										
		1	2	3	4	5	22	23
Actual class	1	23	20	0	3	0	0	0
	2	1	40	1	2	4	4	0
	3	0	3	33	13	2	2	0
	4	0	6	10	34	5	0	0
	5	2	0	0	0	41	0	0

	22	0	0	0	1	0	20	2
	23	0	1	0	0	0	34	1

평가 결과값

- 실제 값 ± 2 범위 내로 예측한 정확도 :

$$\frac{True}{True+False}$$

정확도 : 0.8654019873532068

	value	True	False
0	0	156	3
1	1	223	3
2	2	241	11
3	3	167	16
4	4	91	11
5	5	18	48
6	6	34	19
7	7	18	10
8	8	10	5
9	9	0	10

겨울 x 저녁 x 눈/비 모델의 정확도