Elevating healthcare at Hospital through AI-powered problem solving

Portfolio – 지원자 김로빈

외국계 회사와 협업으로 영어로 작성 된 점 양해 부탁드립니다.

Executive Summary (요약)

- We were asked to predict patient severity at the Emergency Room **(ER use case)** and in-hospital/12-month mortality rates of patients with acute myocardial infraction **(AMI use case)**
- **On ER use case**, we were able to improve the forecast accuracy for **severe cases from the baseline by 211 more critical cases (180% improvement)** with the following highlights:
  - We were able to identify **medically significant insights** from individual experiments but found insights more nuanced in a three-class classification (new target) setup
  - The new target setup is **easier to interpret** and performs better at **prioritising critical cases** than the 3 model setup
  - While the model places emphasis on naïve features, **8 out of 18 top features** contributing to 80% of the model gain (excluding long tail) **were SB features**
  - In totality, **96% (229/244) of all features used in the model** were generated by the platform
  - These **SparkBeyond features contributed to ~32% of total gain** in the model
  - The platform is able to use SHAP values to help physicians understand the model predictions
- **On AMI use case**, **the SparkBeyond platform** validated that the **data was biased against mortality** and tried different data cleaning techniques with minimal results to form a clean data set. Despite that **we were able to identify insights from unbiased columns** that could be of interest to Sejong.
- We propose a **holistic operationalization program** for Sejong hospital addressing **process, platform and people** as next steps.

- 저희는 응급실에서 1.환자의 중증도 (ER 사용 사례)와 2. 급성 심근 경색 환자의 병원 내 사망률/ 12 개월 사망률 (AMI 사용 사례)을 예측하는 과제를 요청 받았습니다.
- ER 사용 사례에서 다음과 같은 밑줄 친 글자들을 통해 Baseline에서 온 심각한 사례에 대한 예측 정확도를 211의 더 많은 중증도 케이스들(180 % 개선)에 의해 개선 할 수 있었습니다.
  - 개별 실험에서 의학적으로 중요한 인사이트를 식별 할 수 있었지만 3개의 클래스 분류 (새롭게 정한 타겟) 설정에서 더 중요한 인사이트를 발견했습니다.
  - 새롭게 만들어진 Target은 3가지 Target 설정보다 해석하기 쉽고 중증도 분류 케이스에 효율적입니다.
  - 모델이 단순한 Feature 들에 중점을 두는 반면 모델 타겟값의 80 %에 기여하는 18 개의 상위 Feature중 8 개 (롱테일 제외)는 Feature였습니다.
  - 전체적으로 모델에 사용 된 모든 Feature의 96 % (229/244)가 플랫폼에서 생성되었습니다.
  - 이러한 SparkBeyond Feature는 모델의 예측률에 32%정도 기여했습니다.
  - 플랫폼은 의사가 모델 예측을 이해하는 데 도움이 되도록 SHAP Value 값을 사용할 수 있습니다.
- AMI 사용 사례에서 플랫폼은 데이터가 사망률에 대해 Biased되어 있는지 확인하고 최소한의 결과로 다양한 데이터 클렌징 기술을 시도하여 깨끗한 데이터 세트를 형성했습니다. 그럼에도 불구하고 우리는 세종의 관심을 끌 수 있는 Unbiased 칼럼에서 인사이트를 확인할 수 있었습니다.
- 다음 단계로 프로세스, 플랫폼 및 사람을 다루는 세종 병원의 전체적인 운영 화 프로그램을 제안합니다.

Recap of context and objectives

Context

- Sejong General Hospital is a private hospital known nationally for it's speciality in Cardiovascular diseases
- The department of Critical care and Emergency Medicine would like to explore the use of Artificial Intelligence based feature discovery to improve healthcare interventions for cardiac and emergency room patients
- Sejong Hospital is looking to improve mortality rates through effective and timely medical interventions for patients

Objectives

- To share with you our approach in tackling the two use cases shared by hospital
- To discuss our findings and perspective on operationalization
- To discuss next steps

Overview

Our understanding of the use cases to explore

Use Cases	Description	What does success look like?
Predicting patient severity at the emergency room	▪ **Classify** end outcomes_(mortality, critical care, hospitalization) of patients based on available data at point of visit to emergency room or after triage	▪ Understand precise drivers of different patient outcomes (e.g. mortality, critical care, hospitalization) ▪ Classify patient outcomes with stable accuracy across different test sets
Predicting mortality risk of patients with acute heart failure	▪ **Classify** in-hospital and 12 month mortality risk of patients based on health indicators	▪ Understand precise drivers of inpatient and 12 month mortality risk ▪ Classify patient outcomes with stable accuracy across different test sets

Expected deliverable – operationalizable predictive model with explainability for selected use case

Status Update

We have proven value in use case 1 and worked around the data issues to provide insights for use case 2

Completed In-progress

	Problem Definition	Data Understanding	Data Preparation	Insights & Validation	Modelling & Evaluation
Predicting patient severity at the emergency room	• SparKanvas aligned, pending confirmation	• Data sources understood and data dictionary formed • Exploratory data analysis performed	• Created target variables • Loaded data into SB platform • Created experiments for interaction variables	• Experimental design • Run hypothesis search in Insights mode • Validate and label top insights	• Experimental design • Build predictive classification models • Hyperparameter tuning • Explanations using SHAP values
Predicting mortality risk of patients with acute heart failure	• SparKanvas aligned, pending confirmation	• Data sources understood • Exploratory data analysis performed	• Managing missing values • Creation of pipelines in SB platform	• Running pipelines for in-hospital mortality • Running pipelines for 12m mortality	

We will come back to Sejong with our perspective on operationalisation

Our approach

Our approach

Our approach rigorously focuses on feature discovery to generate insights to be validated and used in modelling

Heavy Emphasis

	Problem Definition	Data Understanding	Data Preparation	Insights & Validation	Modelling & Evaluation
ACTIVITES	• Define context and objectives • Define target variable and fill SparKanvas	• Sync with Sejong to understand data sources • Confirm time horizon available for each data set • Sync with SB experts on external data	• Create target variables required • Create additional interaction variables required • Load data into SB platform	• Run hypothesis search in Insights mode • Validate and label top insights that contribute and explain the target variable	• Run set of validated insights in SB platform/python in model mode
OUTCOMES	• Target variable aligned • SparKanvas filled	• Clear on variables in each file • Data sources to be used aligned	• Additional features required created and loaded into SB platform	• Validated set of insights for predictive modelling • Decide re-runs with new segment/data	• Predictive model with accuracy validated
HELP NEEDED FROM SEJONG	• Provide perspective on target variable	• Data owner to help with data understanding	• Validate data preparation methods	• Validate insights suggested by the platform	• Provide input on model usability in real world scenarios

Our findings
for ER use case

Defining the problem statement

SparKanvas

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

BUSINESS

Business Challenge:

Establish patient severity classification measures to ensure the right treatment is given to patients at the right time.

Objective (KPI)

What are we trying to achieve and how will it be measured?

Correctly classifying a patient's severity status during triage which corresponds to the final outcomes (death, ICU, hospitalization, outpatient discharge)

- Potential Impact Value: Up to 2% of deaths occur in ER today, exact impact unknown
- KPI Value Today: 2%

Business Drivers

What do we believe has influence on the phenomenon we are analysing?

- Weather conditions
- Emergency room overcrowding
- Time taken to be brought to the hospital
- Average waiting time to be treated

Path to Impact:

How will we use the model / insights to achieve impact? How will this be delivered and / or put into production?

The deep learning-based model is not used in production today due to difficulty of extracting information from EHR system. The plan is to eventually resolve this but need a good baseline model with white-box explanations to augment resource allocation to patients depending on probability of death, ICU etc.

CURRENT STATE

How is this challenge currently tackled, using what technology and are there constraints?

The current approach in all Korean hospitals is use the Korean Triage and Acuity Score (KTAS), a 5-level ranking to score the severity of patient's condition. The appropriate treatment is given based on the severeity score.

TRANSLATION

Target

Classification of Patient Outcomes (Death, ICU, hospitalization, discharge)

How Target relates to KPI

Correctly classifying patients could lead to reduced mortality outcomes

Available Target History

Train Data : 2014. 01. 01 ~ 2016. 06. 30
Test Data : 2018. 09. 01 ~ 2019. 02

Population

Total : 8,981,181 people (Train)
2,604 people (Test)

Evaluation Metric

AUC, NPV, PPV, F1-Score

Current Performance:

Predictive model	AUC [95% CI]	p-value†	p-value‡	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	F1 score [95% CI]
AI + ESI	0.923 [0.920–0.926]	<0.001	–	0.799 [0.795–0.803]	0.857 [0.854–0.859]	0.439 [0.433–0.445]	0.968 [0.966–0.969]	0.567 [0.562–0.571]
AI + KTAS	0.909 [0.906–0.912]	<0.001	<0.001	0.799 [0.795–0.803]	0.859 [0.856–0.862]	0.442 [0.436–0.449]	0.968 [0.906–0.912]	0.569 [0.565–0.575]
AI only	0.867 [0.864–0.871]	–	<0.001	0.799 [0.795–0.803]	0.768 [0.764–0.772]	0.324 [0.318–0.331]	0.965 [0.963–0.968]	0.461 [0.454–0.467]
ESI only	0.839 [0.831–0.846]	<0.001	<0.001	0.357 [0.348–0.365]	0.991 [0.988–0.994]	0.851 [0.843–0.859]	0.917 [0.913–0.921]	0.503 [0.492–0.514]
KTAS only	0.824 [0.815–0.832]	<0.001	<0.001	0.376 [0.368–0.384]	0.971 [0.966–0.975]	0.642 [0.631–0.653]	0.918 [0.912–0.922]	0.474 [0.462–0.486]
NEWS (5)	0.741 [0.734–0.748]	<0.001	<0.001	0.310 [0.299–0.322]	0.976 [0.971–0.980]	0.647 [0.635–0.659]	0.910 [0.902–0.918]	0.419 [0.407–0.431]
MEWS (3)	0.696 [0.691–0.699]	<0.001	<0.001	0.288 [0.275–0.301]	0.936 [0.929–0.942]	0.387 [0.375–0.397]	0.904 [0.898–0.910]	0.330 [0.316–0.343]

DATA

Core Data

-

Internal Data

Korea National Emergency Department Information System (NEDIS) (9,304,887 people)

External Data

DBpedia
Medical Subject Headings
Weather
Google search trends

RECAP: Summary of EDA

Overview of our findings

Dimension	Findings	How will we proceed?
• Data dictionary	• We were able to identify the definition of the columns with NEDIS coding book • Data contained an additional 3 columns compared to the journal papers	• We will run experiments separately on NEDIS and Sejong data to identify critical factors in predicting patient severity
• Sanity check of NEDIS vs Sejong data	• We find 17 additional columns in Sejong data while it does not have a symptom code compared to NEDIS data	
• Sanity check against reference journals	• Number of observations and time periods in data we received differ from data used in journal	• We will procced with data as received
• Duplicate values	• We identify 5,419 rows that exist both in NEDIS and Sejong data (0.01% of NEDIS data set)	• We assume identical values are unique, separate cases and proceed without any deletion of rows
• Target distribution	• Distribution of outcomes across both data sets are similar	• We will test the impact of differing distributions on model performance in the modelling stage
• Patient arrivals	• There appear to be occasional spikes in the number of visitors to ER on certain days	• No action needed, unless directed by Sejong otherwise
• Age distribution	• Sejong data has fewer younger patients with an average age profile roughly 6-7 years older compared to NEDIS	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Gender distribution	• Both data sets contain even distribution of male and female patients	• No action needed
• Responsiveness distribution	• Sejong data has a much lower proportion of non-alert patients on arrival to the ER (1.8%) compared to NEDIS data (5%) • Distribution of outcomes across both responsiveness status in data sets are similar	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Transportation mode	• Severity of patients who arrived by ambulance, police cars and other cars across data seem fairly evenly distributed • Air transport and walking has fewer proportion of severe outcomes in Sejong compared to NEDIS	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Test measurements	• Test results seem to have tight inter-quartiles range with large number of outlier observations • Results from Sejong appear similar in nature to NEDIS, with some differences	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

Summary of EDA

Overview of our findings

Dimension	Findings	How will we proceed?
• Data dictionary	• We were able to identify the definition of the columns with NEDIS coding book • Data contained an additional 3 columns compared to the journal papers	• We will run experiments separately on NEDIS and Sejong data to identify critical factors in predicting patient severity
• Sanity check of NEDIS vs Sejong data	• We find 17 additional columns in Sejong data while it does not have a symptom code compared to NEDIS data	
• Sanity check against reference journals	• Number of observations and time periods in data we received differ from data used in journal	• We will procced with data as received
• Duplicate values	• We identify 5,419 rows that exist both in NEDIS and Sejong data (0.01% of NEDIS data set)	• We assume identical values are unique, separate cases and proceed without any deletion of rows
• Target distribution	• Distribution of outcomes across both data sets are similar	• We will test the impact of differing distributions on model performance in the modelling stage
• Patient arrivals	• There appear to be occasional spikes in the number of visitors to ER on certain days	• No action needed, unless directed by Sejong otherwise
• Age distribution	• Sejong data has fewer younger patients with an average age profile roughly 6-7 years older compared to NEDIS	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Gender distribution	• Both data sets contain even distribution of male and female patients	• No action needed
• Responsiveness distribution	• Sejong data has a much lower proportion of non-alert patients on arrival to the ER (1.8%) compared to NEDIS data (5%) • Distribution of outcomes across both responsiveness status in data sets are similar	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Transportation mode	• Severity of patients who arrived by ambulance, police cars and other cars across data seem fairly evenly distributed • Air transport and walking has fewer proportion of severe outcomes in Sejong compared to NEDIS	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data
• Test measurements	• Test results seem to have tight inter-quartiles range with large number of outlier observations • Results from Sejong appear similar in nature to NEDIS, with some differences	• When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

Data dictionary

Overview and our understanding of field names

Column Name	Korean Definition	English Definition
age	연령	age
onsettovisit	도착 소요 시간	time taken to arrive to ER
sbp	수축기 혈압	systolic blood pressure
dbp	이완기 혈압	diastolic blood pressure
hr	심박수	heart rate
rr	분당 호흡수	respiratory rate
bt	체온	body temperature
sex	성별	sex
ambulance	앰뷸런스 이용여부	transported by ambulance
injury	상해여부	injury or not
spo2	혈중 산소포화도	blood oxygen level
avpu	의식상태의 분류	mental status
PTMIINMN (visit_car)	교통 수단	Mode of transportation
visitdatetime	방문 시간	Visit datetime
onsetdatetime	증상 발현 시간	Onset datetime
crit	중증환자 여부	critical
death	죽음 여부	death
adm	입원 여부	hospitalised

Predictors used for published journals

Additional common columns found in both NEDIS and Sejong data

Target Variables

Sanity check (1/2)

NEDIS vs Sejong Data

NEDIS column	Sejong Column	Korean Definition	English Definition
SYPTCODE	n/a	증상 코드	Symptom code
n/a	main_symptom	주증상1	main symptom
n/a	sub_symptom	주증상2	sub-symptom
n/a	Suicide	자살 여부	suicide
n/a	injury_mech	부상 원인	injury reason
n/a	traffic_accident	교통 사고 여부	Traffic accident
n/a	sit_belt	안전벨트 착용 여부	Sit belt
n/a	child_chair	아동용좌석	Child chair
n/a	front_airbag	전면에어백	Front airbag
n/a	side_airbag	측면에어백	Side airbag
n/a	Helmet	헬멧	helmet
n/a	knee_guide	무릎 및 관전 보호대	Knee or joint protector
n/a	safe_vast	구명 조끼	Safe vast
n/a	no_safe	전혀 착용 안함	No protection equipment
n/a	Notavailable	비해당	Not available
n/a	Noinformation	미상	No information
n/a	main_div	진료과	Department of treatment
n/a	visit_course	내원 경로	Visit course

OBSERVATIONS

- We were able to find more fields in Sejong data
- NEDIS coding database suggests that these additional columns exist but do not appear in the data we received
- The missing symptom code column in Sejong data will be required in order to leverage external data source

Can we get symptom code for Sejong data and missing field in NEDIS data?

Sanity check (2/2)

Overview of train and test data

Category	**Article 1 Reference** Validation of deep-learning-based triage and acuity score using a large national dataset	**Article 2 Reference** Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services	Data we received
Training Data	Source: NEDIS Time period: Jan 2014 – Jun 2016 Rows: 8,981,184	Source: NEDIS Time period: Jan 2014 – Dec 2016 Rows: 8,981,181	Source: NEDIS Time period: Nov 2015 - Dec 2017 Rows: 4,143,234
Test Data	Source: Sejong Time period: Jan-Dec 2017 Rows: 13,989 death: 150 (1.1%) crit: 987 (7.1%) adm: 4,337 (30.0%)	Source: Sejong Time period: Sep 2018 – Feb 2019 Rows: 2,604 death: 30 (1.2%) crit: 319 (12.3%) adm: 1003 (38.5%)	Source: Sejong Time period: Jan 2016-Nov 2019 Rows: 41,140 Time period: Jan 2017-Dec 2017 Rows: 12,452 Time period: Sep 2018 – Feb 2019 Rows: 7,829

Data received differs from the set used in the journal. We will proceed with this data

Duplicate values

Method to identify duplicate values

- We used common columns of NEDIS and Sejong dataset to check for rows with identical sequence of values
- Common columns:
  crit', 'adm', 'death', 'age', 'sex', 'onsettovisit', 'ambulance', 'injury', 'sbp', 'dbp', 'hr', 'rr', 'bt', 'spo2', 'avpu'

Findings

- There are 5,419 rows that exist both in NEDIS and Sejong data
- Considering the size of the data, this is very small proportion, taking only 0.01% of the training data (NEDIS)

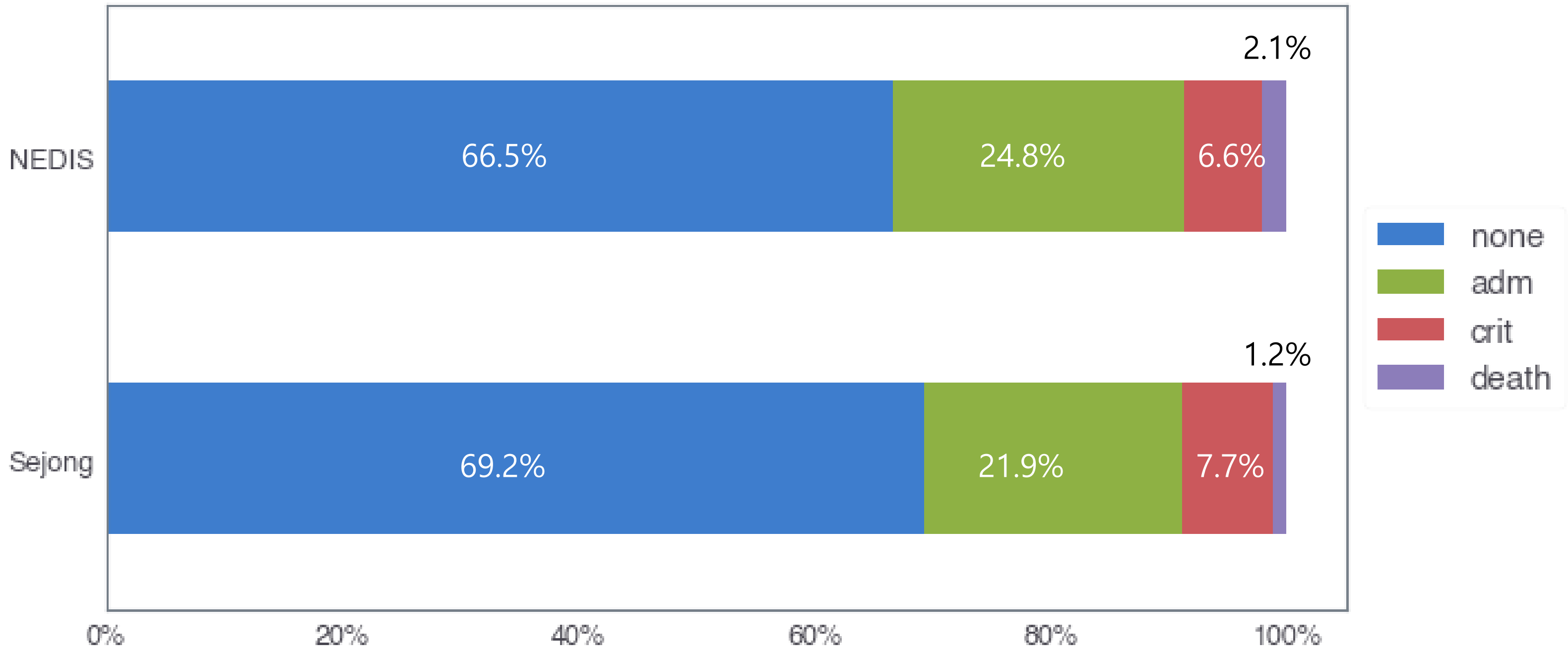
OBSERVATIONS

- Without actual patient identifier, it is not possible to validate the duplicates perfectly
- However, given that the proportion of the identical rows is insignificant, we assume that some values are identical by a coincidence and not duplicates or re-admissions to the ER

We will proceed without removing any duplicate data

Target distribution

Overview of patient outcomes



OBSERVATIONS

- NEDIS and Sejong both have roughly similar target value distribution
- Sejong has higher rate for critical patients while having less proportion of in hospital mortality and hospitalization
- While there are some differences in the target distributions, we believe it to be in acceptable ranges to treat Sejong data as validation data. This assumption will be tested during modelling.

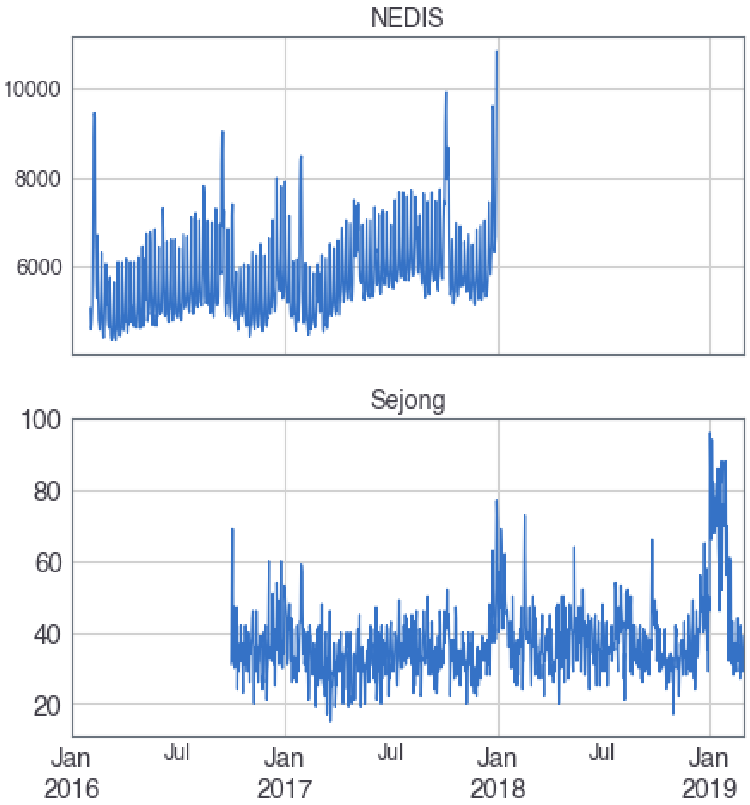
We will test the impact of differing distributions on model performance in the modelling stage

Exploratory Data Analysis

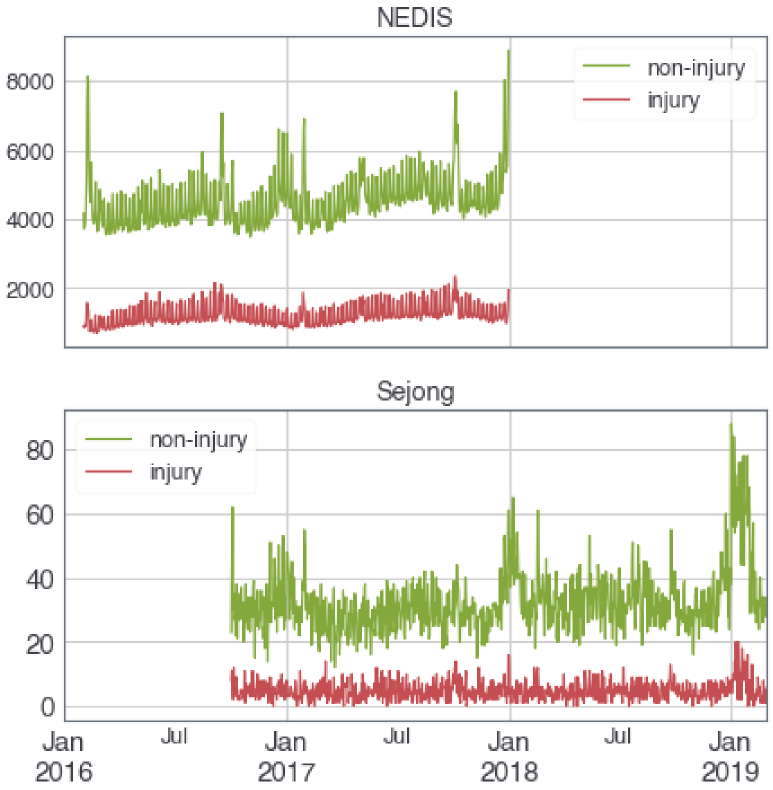
Patient arrivals

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Number of visits to ER over time



Number of visits to ER over time by injury status



OBSERVATIONS

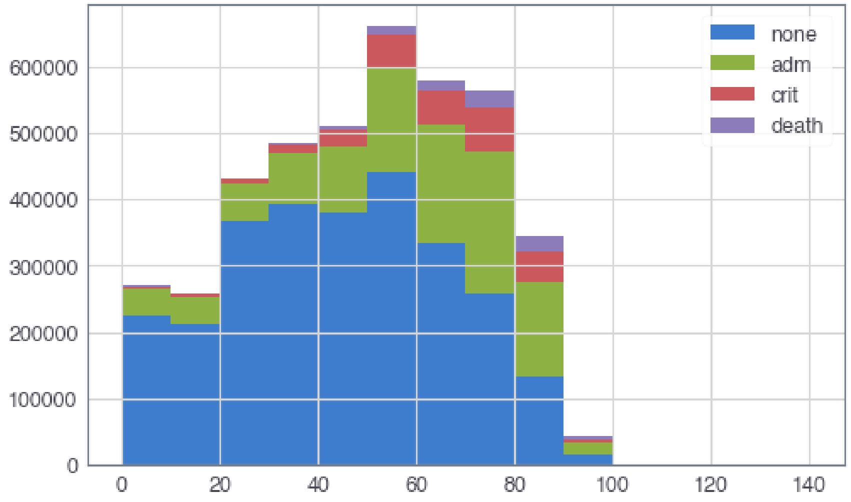
- The time series plots indicates occasional spikes in the number of visitors to ER on certain days
- The spikes are primarily driven by the non-injury cases

Are the time/seasonality trends something we should factor in our models?

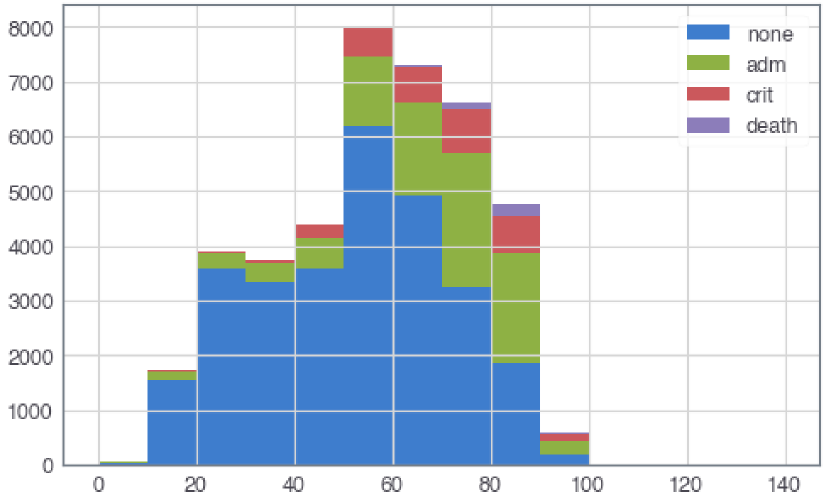
Age distribution

ER patient age profile

NEDIS



Sejong



Percentile	NEDIS	Sejong
25th	31.0	42.0
50th	52.0	58.0
75th	68.0	73.0
mean	49.1	56.1

OBSERVATIONS

- The shape of age distribution is similar between both data sets
- However, Sejong data has fewer younger patients with an average age profile roughly 6-7 years older compared to NEDIS

When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

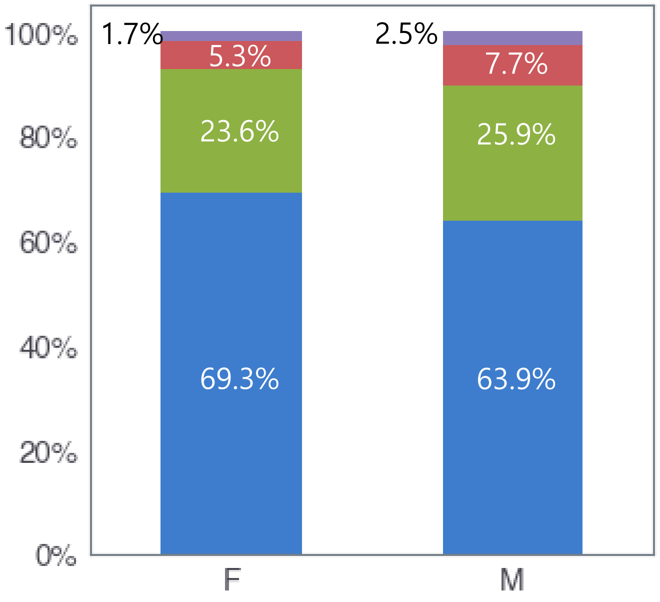
Exploratory Data Analysis

Gender distribution

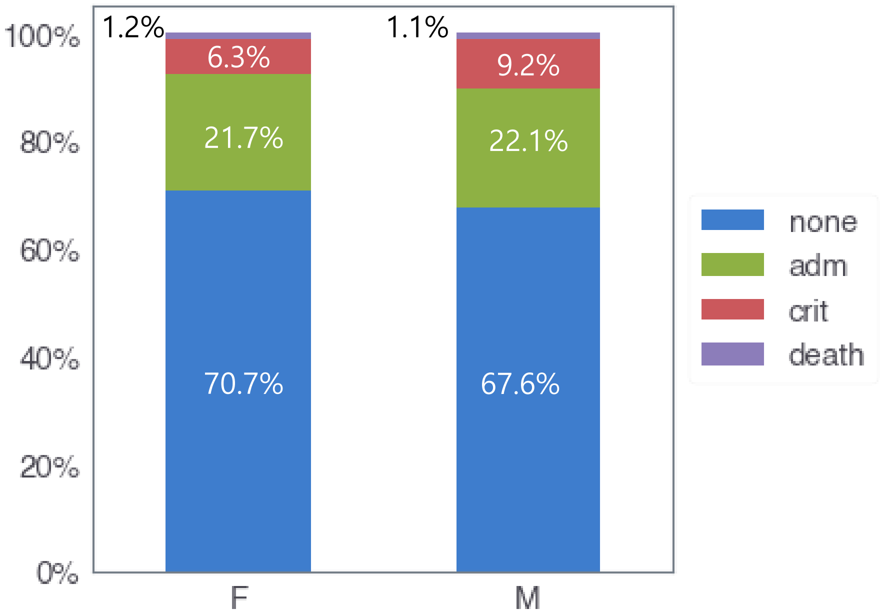
Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

ER patient gender profile

NEDIS



Sejong



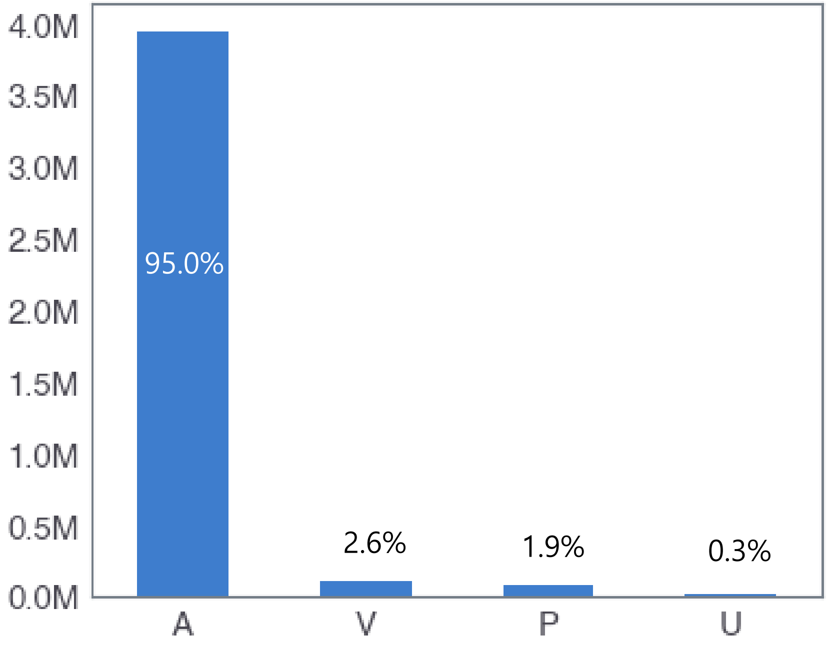
OBSERVATIONS

- In both NEDIS and Sejong, males have a higher probability of severe outcomes
- Males however seem to be have ~4% less chance of severee outcomes in Sejong compared to NEDIS

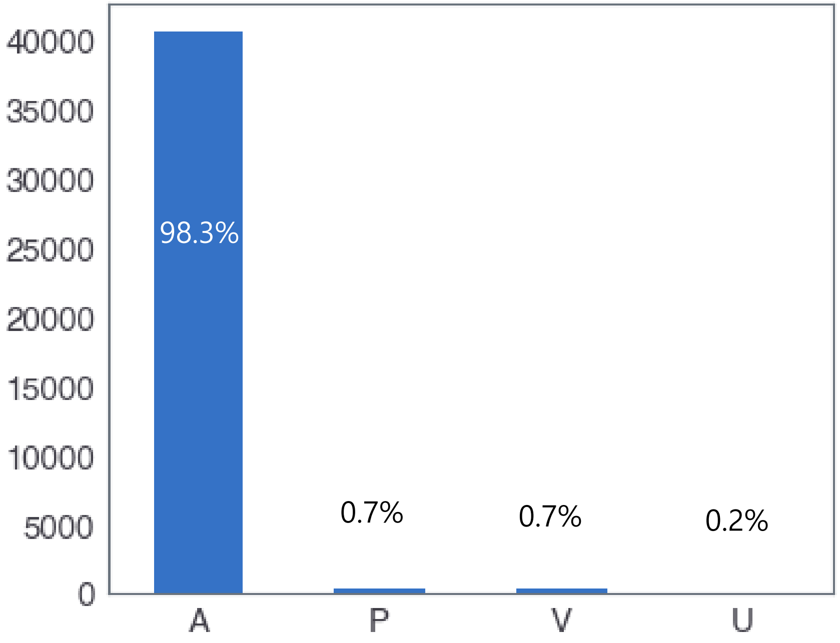
Data sets contain even distribution of male and female patients

Responsiveness distribution (1/2)

ER patient responsiveness status (NEDIS)



ER patient responsiveness status (Sejong)



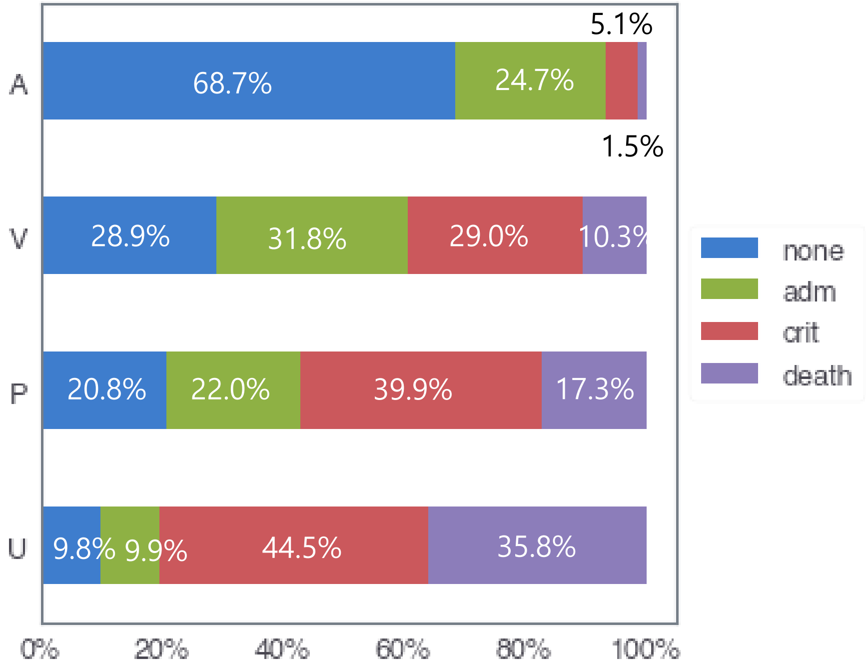
OBSERVATIONS

- Sejong data has a much lower proportion of alter patients on arrival to the ER (1.8%) compared to NEDIS data (5%)

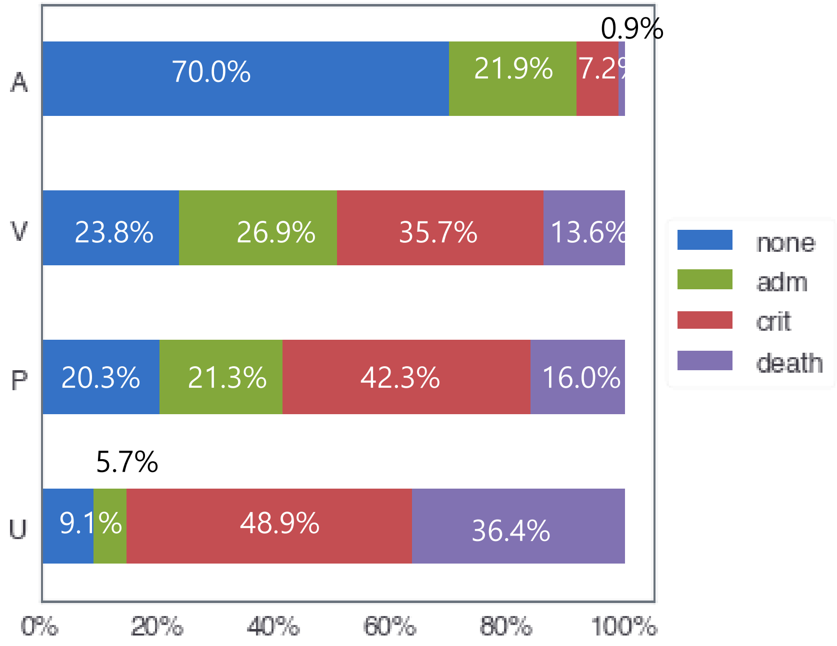
When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

Responsiveness distribution (2/2)

ER patient outcomes by responsiveness status (NEDIS)



ER patient outcomes by responsiveness status (Sejong)



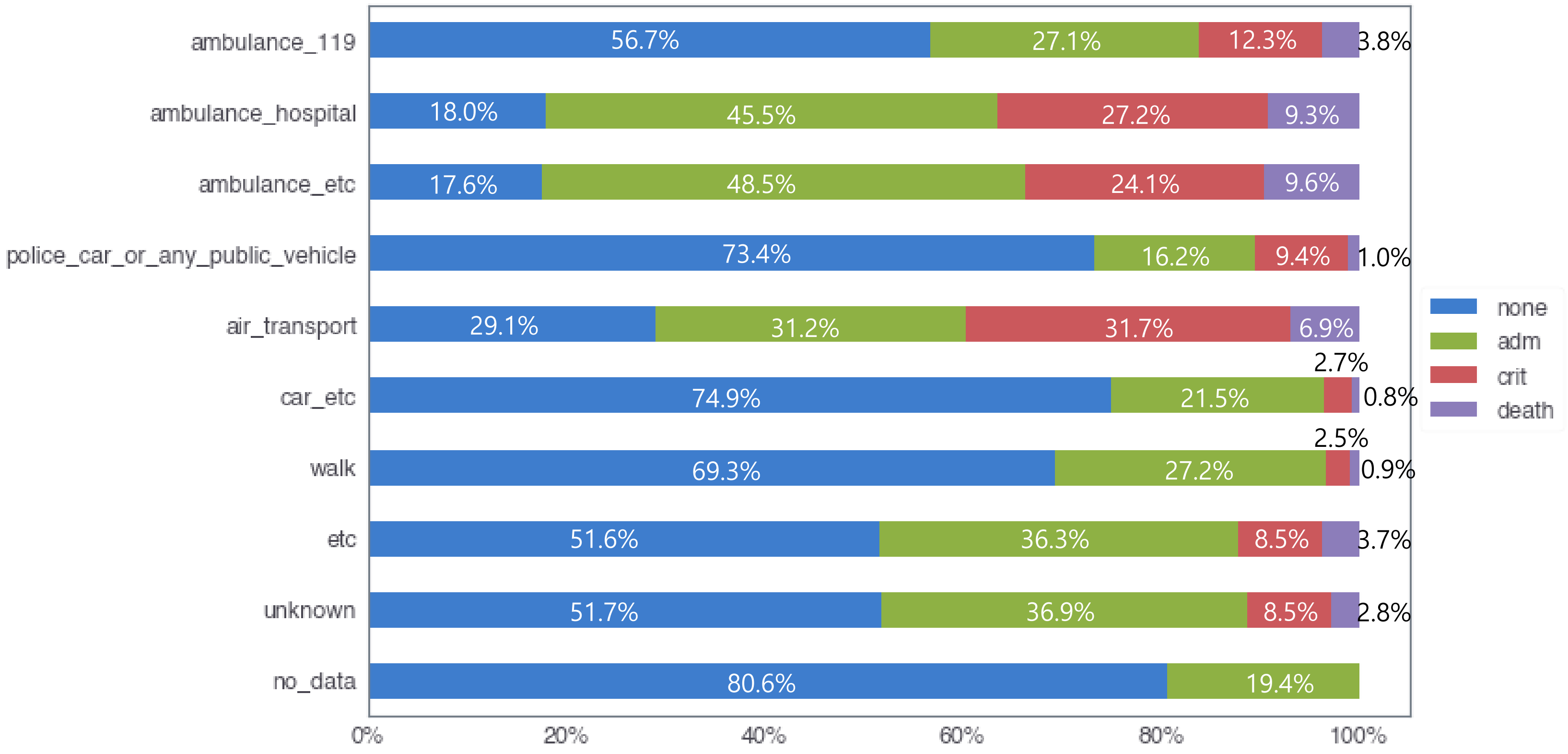
OBSERVATIONS

- Mortality was lowest when patients came in responsive
- Patients who came in unresponsive were more likely to have an undesirable end outcome as compared to patients who were responsive

Distribution of outcomes across both responsiveness status in data sets are similar

Transportation mode (1/2)

ER patient outcomes by mode of transportation (NEDIS)



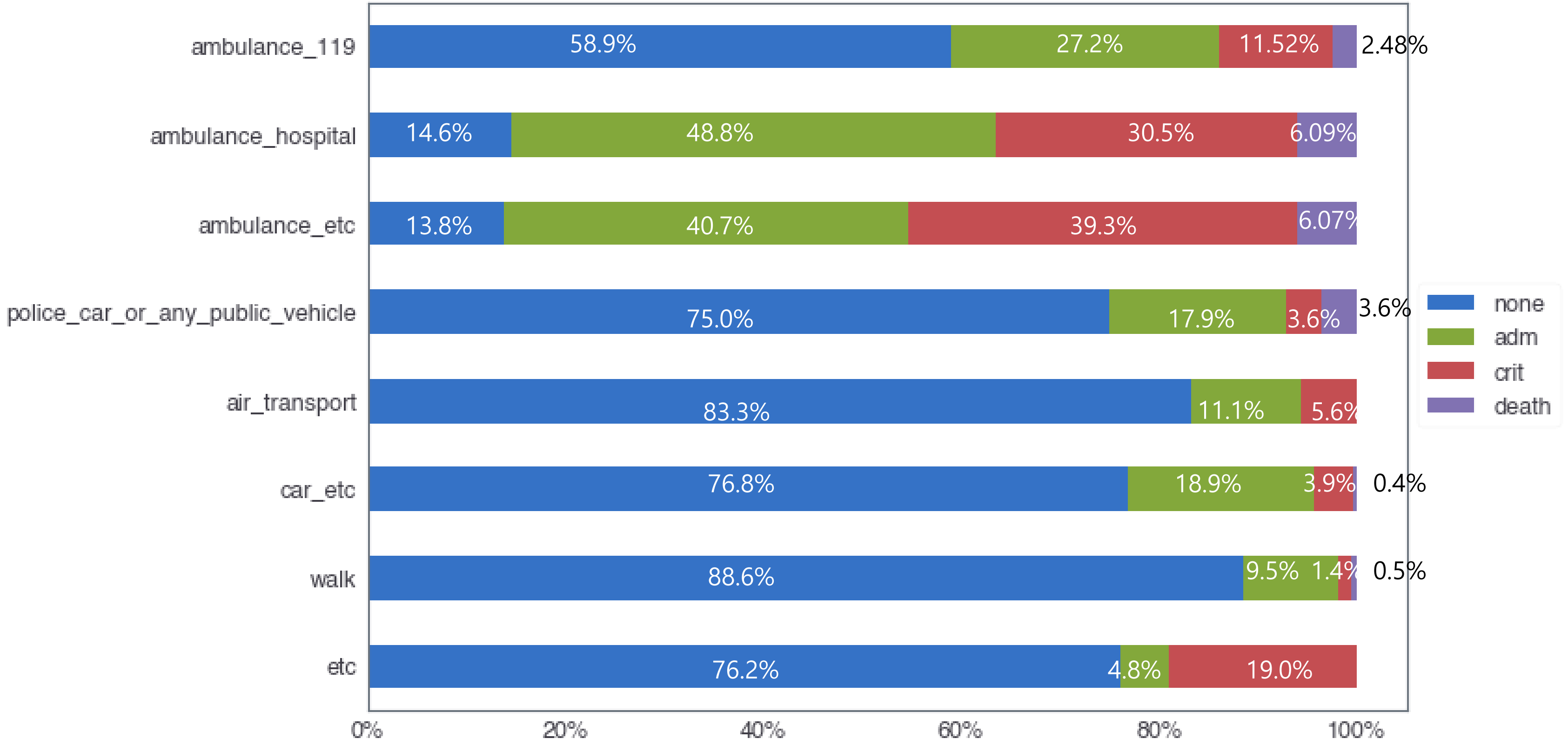
OBSERVATIONS

- Patients transported via ambulance or aircraft are more likely to have severe outcomes
- Depending on the type of ambulance used, the severity rate differs
- While the mode of transportation reflects the urgency of a patient, this is merely a correlation

Mode of transport could be an important correlator to outcomes, though no causal link can necessarily be established

Transportation mode (2/2)

ER patient outcomes by mode of transportation (Sejong)



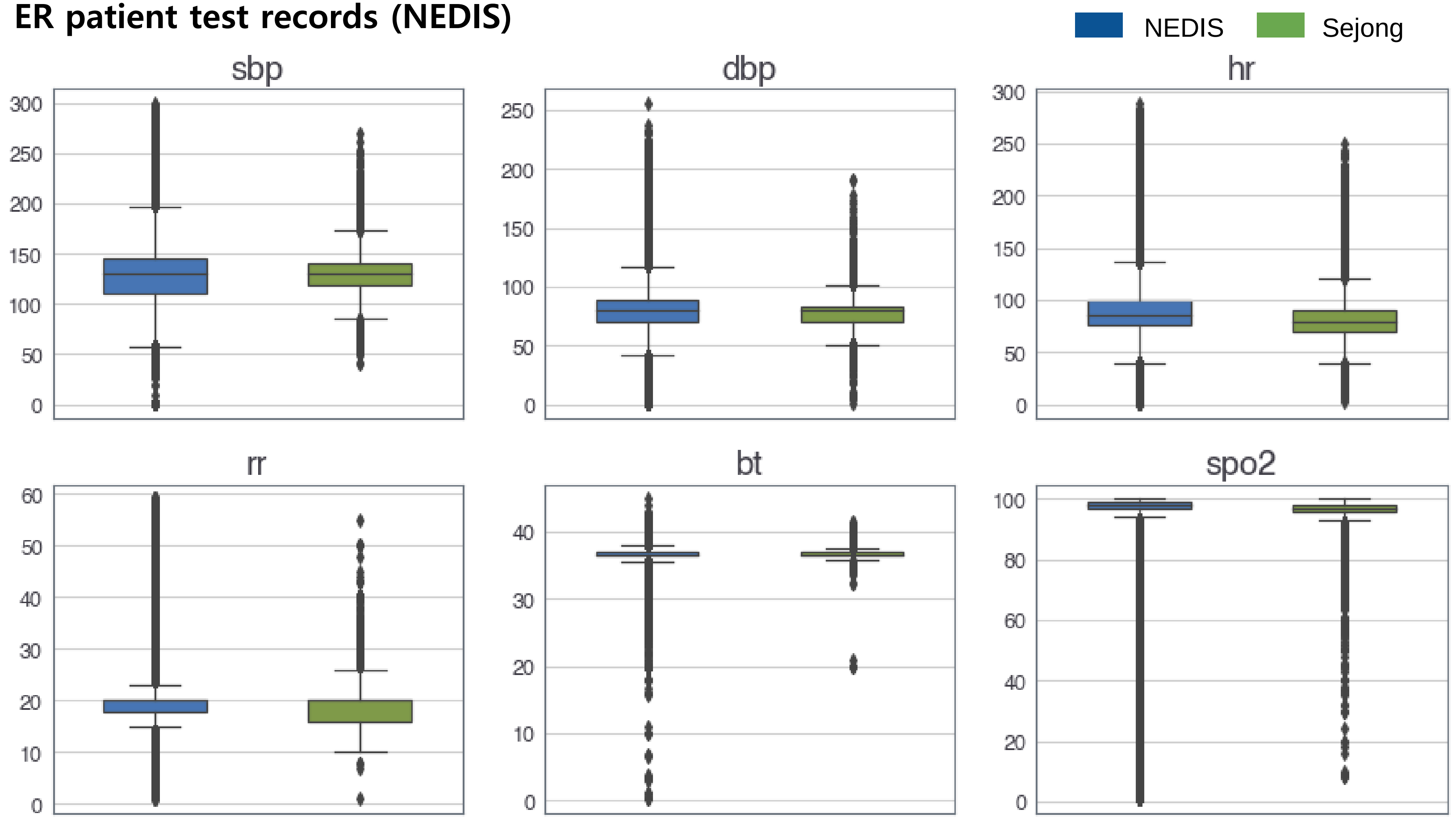
OBSERVATIONS

- Severity of patients who arrived by ambulance, police cars and other cars across data seem fairly evenly distributed
- Air transport and walking has fewer proportion of severe outcomes in Sejong compared to NEDIS

When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

Test Measurements

ER patient test records (NEDIS)



OBSERVATION

- Test results seem to have tight inter-quartile range with large number of outlier observations across both NEDIS and Sejong
- Heart rate, respiratory rate, and oxygen level in blood show slight variation between NEDIS and Sejong

When modelling, we may take additional cuts of train/test splits to ensure even distribution of data

External data

External data sources explored

Description	Source	Time Horizon	Granularity	Status	Comment
Weather	Korea national	2014-2019	Daily	Done	• We were able to connect weather data to the modelling data • However, it is hard to establish the causality between weather and individual fatality
UMLS subject text to DBpedia mapping	http://denote.rnet.ryerson.ca/umlsMap/	N/A	Symptom code	Done	• Medical subject headings requires symptom code, which is missing in Sejong data • In the meantime, we will proceed with NEDIS data for experiments
Google search trends				Pending	• We need guidance on keywords we should search for in Google search trends
DBPedia	DBpedia	N/A	N/A	Done	• DBpedia is a crowd-sourced community effort to extract structured information from DBpedia and make this information available on the Web • Words in sxcode look-up table will be searched in DBPedia

QUESTIONS

- What are the keywords we should search for in Google search trends?
- What are potential other data sources to explore?

Target Variable Treatment

RECAP: Data preparation

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Overview of target variables to be tested

Current Approach – treat each outcome individually

adm	crit	death
1	1	1
1	1	0
1	1	1
1	0	0
1	0	0
0	0	0

To be tested – combine outcomes

new target
Critical
Critical
Critical
Hospitalization
Hospitalization
None

COMMENTARY

- We adopt building 3 different models for each class in line with Sejong's current approach
- For future iterations, we propose to merge the target variable into a modified critical target
- However, we will need to understand how each class is determined and verify whether there's no false ground truth e.g. (hospitalisation=1, crit=0 death=1)

Will this new target variable be helpful for medical practitioners?

RECAP: Experimental Design

Overview of experimental runs

No	Description	Data Set	Target Variables	Conclusion
1	Only columns used in the journal papers	NEDIS	Death Critical Admission New target	Most information available at point of triage is insightful for predicting severe cases
2	Additional columns found in Sejong data set	Sejong		We find specific insights related to Sejong patients that not appear in NEDIS but not related to the additional columns found in Sejong data
3	With symptom code and mode of transportation	NEDIS		Interaction of triage information and symptom codes are insightful to predict severe cases
4	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS		Symptoms translated to related topics in English can help distinguish cases with higher and lower risk of severe cases
5	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS		Symptoms translated to related medical topics in English can help distinguish cases with higher and lower risk of severe cases
6	With google search trends	NEDIS		Shelved for future work

KEY TAKEAWAYS

- We find that data from training set provides insights in general about severe cases
- Linking up external data sources provides additional insights into the type of conditions that are related severe cases
- When running experiments of the target variable separately, the insights generated **tend not to be very different as opposed to the new target which could differentiate amongst outcomes**
- Testing for under sampling of non-severe cases (since death rates is very low) and crowd levels at ER yielded no additional insights

We have a good number of features to start modelling

Insights Experiments

Experimental Design and Outcomes

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Overview of experimental runs

No	Description	Data Set	Target Variables	Conclusion
1	Only columns used in the journal papers	NEDIS	Death Critical Admission New target	Most information available at point of triage is insightful for predicting severe cases
2	Additional columns found in Sejong data set	Sejong		We find specific insights related to Sejong patients that not appear in NEDIS but not related to the additional columns found in Sejong data
3	With symptom code and mode of transportation	NEDIS		Interaction of triage information and symptom codes are insightful to predict severe cases
4	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS		Symptoms translated to related topics in English can help distinguish cases with higher and lower risk of severe cases
5	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS		Symptoms translated to related medical topics in English can help distinguish cases with higher and lower risk of severe cases
6	With google search trends	NEDIS		Shelved for future work

KEY TAKEAWAYS

- We find that data from training set provides insights in general about severe cases
- Linking up external data sources provides additional insights into the type of conditions that are related severe cases
- When running experiments of the target variable separately, the insights generated **tend not to be very different as opposed to the new target which could differentiate amongst outcomes**
- Testing for under sampling of non-severe cases (since death rates is very low) and crowd levels at ER yielded no additional insights

We have a good number of features to start modelling

Emerging insights: death (1/3)

Overview of experimental runs

No	Target	Description	Data Set	Number of Insights	Insightful	Conclusion	Link
1	Death	Only columns used in the journal papers	NEDIS	15	13	Most information available at point of triage is insightful for predicting death	http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/17
2	Death	Additional columns found in Sejong data set	Sejong	131	10	Shortness of breath and chronic weakness are specifically helpful to predict deaths for Sejong patients	http://52.141.17.15/#/visualPipeline/SJ_categorical/23
3	Death	With symptom code and mode of transportation	NEDIS	191	15	Interaction of triage information and symptom codes are insightful to predict death	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed/1
4	Death	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS	269	16	Symptoms translated to related topics in English can help distinguish cases with higher and lower risk of death	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v2_external_data/17
5	Death	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS	62	25	Symptoms translated to related medical topics in English can help distinguish cases with higher and lower risk of death	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v3_external_data/12
6	Death	With google search trends	NEDIS			Pending	

Insights Experiments

Emerging insights: death (2/3)

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Overview of insights

No	Insights related to Death		Insights related to Survival
1	• Low blood oxygen • Non-alert patients • Older patients • Arrival by ambulance • Low systolic blood pressure	• High heart rate • Higher respiratory rate • Low diastolic blood pressure • Abnormal body temperature • Time of arrival to ER from symptom onset > 2 days	• Being female • Patients with injury
2	• Main division serving patient is ER • Severe shortness of breath • Mild shortness of breath • Chronic weakness • Older patients (>70)	• Oxygen level between 43 and 94 • Low systolic BP • Diastolic BP outside of 70 and 99 • High heart rate • Non infectious disease	
3	• Higher age with lower oxygen level • Higher age with higher heart rate • Higher age with lower systolic BP • Higher age with higher respiratory rate • Lower oxygen level with higher respiratory rate	• Lower oxygen level with higher heart rate • Arrival by ambulance • Ambulance took more than an hour to reach hospital • Symptoms with Dyspnoea, Asthenia	
4	• Symptoms related to psychological feature, mental, connective tissue, animal tissue, nasal, dyspena, high frequency of the use of word pathologic • Symptoms not related to protected daisy		• Symptoms related to dizziness, skin eruption, head and neck, skeletal system, footwear, upper limb, care and hygiene, dermatology, basic cognitive process, upper respiratory tract
5	• Symptoms related to dyspnea, mental orientation, mental status, mental health, blood, lung		• Symptoms related to dizziness, rash, wound, human eye, sense, headache, ankle, chest pain, abdominal pain, pain, psychosis, pharynx, finger, knee, human nose, foreign body, hand, ear, foot

Emerging insights: death (3/3)

Summary statistics of selected features

Description	Patients who died (2.1%) NEDIS	All patients (100%) NEDIS	Patients who died (1.2%) Sejong	All patients (100%) Sejong
Age (average)	71.1	49.1	76.6	56.1
Sex (% Female)	39.7%	48.7%	53.3%	52.2%
Injury rate (average)	7.1%	21.2%	4.9%	13.2%
Oxygen level (average)	92.1	97.4	91.9	96.6
Heart Rate (average)	99.5	89.5	92.6	82.7
Respiratory Rate (average)	22.0	20.1	21.7	19.0
Systolic BP (average)	120.1	130.1	123.8	129.5
Diastolic BP (average)	71.6	78.0	73.6	78.2
Body Temperature (average)	36.8	36.8	36.7	36.7
Ambulance arrival rate (%)	72.7%	29.2%	74.3%	26.4%
Time of onset to ER arrival (average)	56.9	41.0	35.5	29.5
Time taken for ambulance to reach hospital (average hours)	45.1	27.8	27.3	21.8
Top 5 Symptom Codes	C0000737: abdominal pain C0004093: Asthenia C0013404: dyspnea C0015967: Fever C0233407: Disorientation	C0000737: abdominal pain C0012833: dizziness C0013404: dyspnea C0015967: Fever C0018681: Headache	N/A	N/A
Related keywords from Symptom Codes and DBpedia	Dyspnea, Mental health, Blood, Lung		N/A	N/A

Emerging insights: critical (1/3)

Overview of experimental runs

No	Target	Description	Data Set	Number of Insights	Insightful	Synthesis	Link
1	Critical	Only columns used in the journal papers	NEDIS	14	10	Most information available at point of triage is insightful for predicting criticality	http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/18
2	Critical	Additional columns found in Sejong data set	Sejong	157	10	Mild or severe pain, no-infection, experience of doctor (?) and age is important for critical patients in Sejong	http://52.141.17.15/#/visualPipeline/SJ_categorical/24
3	Critical	With symptom code and mode of transportation	NEDIS	178	13	Interaction of triage information and symptom codes are insightful to predict criticality	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed/4
4	Critical	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS	286	15	Symptoms translated to related topics in English can help distinguish between higher and lower risk of critical cases	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v2_external_data/16
5	Critical	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS	61	28	Symptoms translated to related medical topics in English can help distinguish cases with higher and lower risk of death	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v3_external_data/13
6	Critical	With google search trends	NEDIS			Pending	

Emerging insights: critical (2/3)

Overview of insights

No	Insights related to Critical		Insights related to Admission/Discharge
1	• Low blood oxygen • Non-alert patients • Arrival by ambulance • Low systolic blood pressure	• Higher heart rate • Higher respiratory rate • Low diastolic blood pressure • Abnormal body temperature	• Being female • Patients with injury
2	• Main division serving patient is ER • Internal medicine division • Mild or severe pain, no-infection • Junior doctor (?) • Older patients (>62)	• Lower oxygen level • Diastolic BP outside of 60 and 99 • Higher heart rate • Higher respiratory • Arrival by ambulance	
3	• Higher aged patient who arrive unalert • Ambulance arrival with lower oxygen level • Higher age with higher heart rate • Higher age with lower systolic BP • Higher age with higher respiratory rate • Lower oxygen level with higher respiratory rate	• Lower oxygen level with higher heart rate • Arrival by ambulance • Symptoms with mental or behavioural dysfunction	• Female aged <75 • Arrival by care and age <50 • Symptoms with abdominal pain
4	• Symptoms related to psychological feature, vertebrate, toxicology, respiratory system, dyspena, high frequency of the use of word pathologic, neck pain • Symptoms not related to cognition		• Symptoms related to protected daisy, head and neck, skeletal system, upper limb, care and hygiene, basic cognitive process, digestive system
5	• Symptoms related to dyspnea, weakness, hemiparesis, mental orientation, mental status, mental health, blood pressure, substance intoxication, cardiac arrest, muscle weakness		• Symptoms related to dizziness, rash, pain abdominal pain, pain, wound, sense, human eye, dizziness, arthralgia, human nose, skin, foreign body, pharynx, finger, human gastrointestinal tract, fever, hand, foot

Emerging insights: critical (3/3)

Summary statistics of selected features

Description	Patients in critical condition (8.7%) NEDIS	All patients (100%) NEDIS	Patients in critical condition (8.9%) Sejong	All patients (100%) Sejong
Age (average)	65.3	49.1	68.6	56.1
Sex (% Female)	39.8%	48.7%	44.5%	52.2%
Injury rate (average)	15.4%	21.2%	3.7%	13.2%
Oxygen level (average)	94.4	97.4	93.6	96.6
Heart Rate (average)	93.3	89.5	87.8	82.7
Respiratory Rate (average)	21.2	20.1	20.5	19.0
Systolic BP (average)	129.0	130.1	130.4	129.5
Diastolic BP (average)	76.3	78.0	77.4	78.2
Body Temperature (average)	36.7	36.8	36.7	36.7
Ambulance arrival rate (%)	70.5%	29.2%	64.7%	26.4%
Time of onset to ER arrival (average)	38.4	41.0	31.7	29.5
Time taken for ambulance to reach hospital (average hours)				
Top 5 Symptom Codes	C0000737: abdominal pain C0008031: Mastodynia C0013404: dyspnea C0015967: Fever C0233407: Disorientation	C0000737: abdominal pain C0012833: dizziness C0013404: dyspnea C0015967: Fever C0018681: Headache	N/A	N/A
Related keywords from Symptom Codes and DBpedia	Blood pressure, weakness, Substance intoxication, Cardiac arrest		N/A	N/A

Emerging insights: admission (1/3)

Overview of experimental runs

No	Target	Description	Data Set	Number of Insights	Insightful	Conclusion	Link
1	Admission	Only columns used in the journal papers	NEDIS	15	14	Most information available at point of triage is insightful for predicting admission	http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/19
2	Admission	Additional columns found in Sejong data set	Sejong	64	12	No new information	http://52.141.17.15/#/visualPipeline/SJ_categorical/25
3	Admission	With symptom code and mode of transportation	NEDIS	162	18	Interaction of triage information and symptom codes are insightful to predict admission	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed/9
4	Admission	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS	290	11	Symptoms translated to related topics in English can help distinguish cases between higher and lower risk of admission	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v2_external_data/21
5	Admission	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS	61	25	Symptoms translated to related to medical topics in English can help distinguish cases between higher and lower risk of admission	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v3_external_data/14
6	Admission	With google search trends	NEDIS			Pending	

Insights Experiments

Emerging insights: admission (2/3)

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Overview of insights

No	Insights related to Admission		Insights related to Discharge
1	• Older patients • Low blood oxygen • Non-alert patients • Arrival by ambulance • Low systolic blood pressure	• Abnormal heart rate • Higher respiratory rate • Low diastolic blood pressure • Higher body temperature • Time of arrival to ER from symptom onset > 15hrs	• Being female • Patients with injury
2	• Main division serving patient is ER • Internal medicine division • Older patients (>63) • Arrival by ambulance	• Lower oxygen level • Lower Diastolic and Systolic BP • Abnormal heart rate • Lower respiratory rate	• Symptom contains ache or fever (non-infection)
3	• Higher aged patient who arrive unalert • Ambulance arrival with onset > 1.5hr • Ambulance arrival with older patient • Higher age with higher heart rate	• Higher age with higher respiratory rate • Higher oxygen level with higher age • Symptoms with dyspnea, dysarthria, haematochezia, icterus, mental or behavioural dysfunction	• Non ambulance arrival with onset <10hrs • Non ambulance arrival with injury • Symptoms related to exanthema, itching, dizziness, epistaxis
4	• Symptoms related to pathophysiology, inhalation • Symptoms without reference to body parts		• Symptoms related to protected daisy, head and neck, injury, facial features, skin care, skeletal system, upper respiratory tract
5	• Symptoms related to laboured breathing, weakness, blood, muscle weakness, yellow, blood pressure, abnormal behaviour, hemiparesis, childbirth, hip, fever		• Symptoms related to rash, human eye, wound human nose, psychosis, foreign body, skin, pharynx, finger, respiratory rate, ear, tooth, bee, middle east

Emerging insights: admission (3/3)

Summary statistics of selected features

Description	Patients Admitted (33.5%) NEDIS	All patients (100%) NEDIS	Patients Admitted (30.8%) Sejong	All patients (100%) Sejong
Age (average)	59.6	49.1	67.3	56.1
Sex (% Female)	44.7%	48.7%	49.7%	52.2%
Injury rate (average)	14.3%	21.2%	7.2%	13.2%
Oxygen level (average)	96.2	97.4	95.3	96.6
Heart Rate (average)	91.7	89.5	84.7	82.7
Respiratory Rate (average)	20.5	20.1	19.6	19.0
Systolic BP (average)	130.4	130.1	131.3	129.5
Diastolic BP (average)	77.2	78.0	78.0	78.2
Body Temperature (average)	36.9	36.8	36.8	36.7
Ambulance arrival rate (%)	46.6%	29.2%	45.0%	26.4%
Time of onset to ER arrival (average)	54.8	41.0	40.1	29.5
Time taken for ambulance to reach hospital (average hours)				
Top 5 Symptom Codes	C0000737: abdominal pain C0004093: Asthenia C0008031: Mastodynia C0013404: dyspnea C0015967: Fever	C0000737: abdominal pain C0012833: dizziness C0013404: dyspnea C0015967: Fever C0018681: Headache	N/A	N/A
Related keywords from Symptom Codes and DBpedia	Laboured breathing, Childbirth, Hip, Fever		N/A	N/A

Emerging insights: new target (1/3)

Overview of experimental runs

No	Target	Description	Data Set	Number of Insights	Insightful	Conclusion	Link
1	Modified Critical	Only columns used in the journal papers	NEDIS	15	12	Most information available at point of triage is insightful for predicting criticality	http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/21
2	Modified Critical	Additional columns found in Sejong data set	Sejong	95	17	New information on shortness of breath and breast pain	http://52.141.17.15/#/visualPipeline/SJ_categorical/40
3	Modified Critical	With symptom code and mode of transportation	NEDIS	222	26	Interaction of triage information and symptom codes are insightful to predict admission	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed/10
4	Modified Critical	With symptom code, mode of transportation and DBpedia (general symptom keyword search on DBpedia)	NEDIS	300	20	Symptoms translated to related topics in English can help distinguish cases between higher and lower risk of admission and critical	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v2_external_data/21
5	Modified Critical	With symptom code, mode of transportation, DBpedia and UMLS (UMLS keyword search on DBpedia)	NEDIS	66	29	Symptoms translated to related medical topics in English can help distinguish cases with higher and lower risk of admission and critical	http://52.141.17.15/#/visualPipeline/Usecase1_ER_preprocessed_v3_external_data/19
6	Modified Critical	With google search trends	NEDIS			Pending	

Emerging insights: new target (2/3)

Overview of insights

No	Insights related to Critical	Insights related to Admission	Insights related to Discharge
1	• Arrival by ambulance • Higher age • Lower blood oxygen • Non-alert patient • Onset <5 hours • Low systolic/diastolic blood pressure • Higher heart rate • Higher respiratory rate	• Higher body temperature	• Patients with injury • Female patients
2	• Treatment in other division • Arrival by ambulance • Non infectious disease • Symptoms related to dyspnea, breath/mild breathing problem, breast pain	• Older doctor (?) • Main division serving patient is ER • Onset time >8hrs from time to visit ER • Same day ER visit as onset • Higher body temperature • Symptom related to slippage	• Fever • Palpitation
3	• Higher respiratory rate, with higher age • Higher heart rate, with higher age • Lower systolic BP, with higher age • Symptoms related to dyspnea, mastodynia, mental or behavioural dysfunction, pulse rate	• Higher body temperature, higher age • Same day ER visit as symptom onset, abdominal pain, icterus, haematochezia, hip pain	• Arrive by car and age <52 • Onset < 5hrs and age <72 • Not arriving by ambulance and having injury • Arrival after 6pm or weekend • Symptom with Exanthema, dizziness, injury or poisoning, epistaxis, preventative procedure
4	• Symptom with pathophysiology, protected daisy, vertebrate, drug abuse, toxicology, breathing, motor weakness, inhalation • Symptom does not contain body part, cognition	• Symptom contains blood in stool	• Symptom related to sensory organs, head and neck, injury, skeletal system, pain, upper respiratory tract, upper limb anatomy, care and hygiene, facial features
5	• Symptoms related to dyspnea, muscle weakness, weakness, blood pressure, blood, abnormal behaviour, speech, toxicity, mental orientation	• Symptoms related to fever, abdominal pain, sense, obstetrics, abdomen	• Symptoms related to rash, human eye, psychosis, wound, human nose, foreign body, pain, skin, pharynx, finger, ear, middle east, tooth, hand

Emerging insights: new target (3/3)

Summary statistics of selected features

Description	Critical (8.7%) NEDIS	Admitted (24.8%) NEDIS	All patients (100%) NEDIS	Critical (8.9%) Sejong	Admitted (21.9%) Sejong	All patients (100%) Sejong
Age (average)	65.3	59.6	49.1	68.6	67.3	56.1
Sex (% Female)	39.8%	44.7%	48.7%	44.5%	49.7%	52.2%
Injury rate (average)	15.4%	14.3%	21.2%	3.7%	7.2%	13.2%
Oxygen level (average)	94.4	96.2	97.4	93.6	95.3	96.6
Heart Rate (average)	93.3	91.7	89.5	87.8	84.7	82.7
Respiratory Rate (average)	21.2	20.5	20.1	20.5	19.6	19.0
Systolic BP (average)	129.0	130.4	130.1	130.4	131.3	129.5
Diastolic BP (average)	76.3	77.2	78.0	77.4	78.0	78.2
Body Temperature (average)	36.7	36.9	36.8	36.7	36.8	36.7
Ambulance arrival rate (%)	70.5%	46.6%	29.2%	64.7%	45.0%	26.4%
Time of onset to ER arrival (average)	38.4	54.8	41.0	31.7	40.1	29.5
Time taken for ambulance to reach hospital (average hours)						
Top 5 Symptom Codes	C0000737: abdominal pain C0008031: Mastodynia C0013404: dyspnea C0015967: Fever C0233407: Disorientation	C0000737: abdominal pain C0004093: Asthenia C0008031: Mastodynia C0013404: dyspnea C0015967: Fever	C0000737: abdominal pain C0012833: dizziness C0013404: dyspnea C0015967: Fever C0018681: Headache	N/A	N/A	N/A
Related keywords from Symptom Codes and DBpedia	Dyspnea, Muscle weakness, Blood pressure, Blood	Fever, Abdominal pain, Obstetrics, Sense		N/A	N/A	N/A

Experimental Design

Overview of experimental design

Iteration 1: Identifying best model

- Using various modelling techniques, we identify the best performing configuration and model for each target variable
- This allows us to compare performance against a baseline experiment to test for improvements

Iteration 2: Testing for internal consistency

- We check for internal consistency by understanding how many predictions across the 3 models give results which are not possible in the training set (e.g. death but not admitted)
- This allows us to test whether operationalizing 3 models would sufficiently produce consistent results

Iteration 3: Comparing against new target

- We compare the outcomes of using 3 models against a single model using a three-class target variable
- This tells us which of these setups would be best for Sejong hospital when operationalizing the model(s)

Iteration 1

Modelling Design

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Overview of modelling runs

Non-Critical Samples | Critical Samples | Discarded

Baseline | **Re-Sampling** | **Down-Sampling + Re-Sampling** | **Sejong**

DATA SPLIT

NEDIS Data

Sejong Data

	Baseline	Re-Sampling	Down-Sampling + Re-Sampling	Sejong
DESCRIPTION	• Use NEDIS as train and Sejong as test data in line with methodology used in reference journals	• Re-sample training data (NEDIS) to match distribution of **age group** in test data (Sejong)	• Down-sample **deaths** from BOTH data sets to balance against **non-death cases**	• Use Sejong data as train **(80%)** and test **(20%)** to test for differences in learning parameters
WHY DO WE DO IT?	• Establish a reference point for performance	• Prevent the model from over-learning patterns in the data that could be less (frequently) relevant to Sejong hospital	• Model accuracy could be heavily driven by non critical cases, which are of less importance to doctors than the critical cases	• Establish understanding if learning from NEDIS data could be an disadvantage for Sejong hospital

PARAMETERS

- We ran each experiment with features used in reference journal and SparkBeyond generated features
- We explored hyperparameters tuning packages to tune and select models

Iteration 1

Modelling Outcomes (1/2)

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | **Modelling & Evaluation**

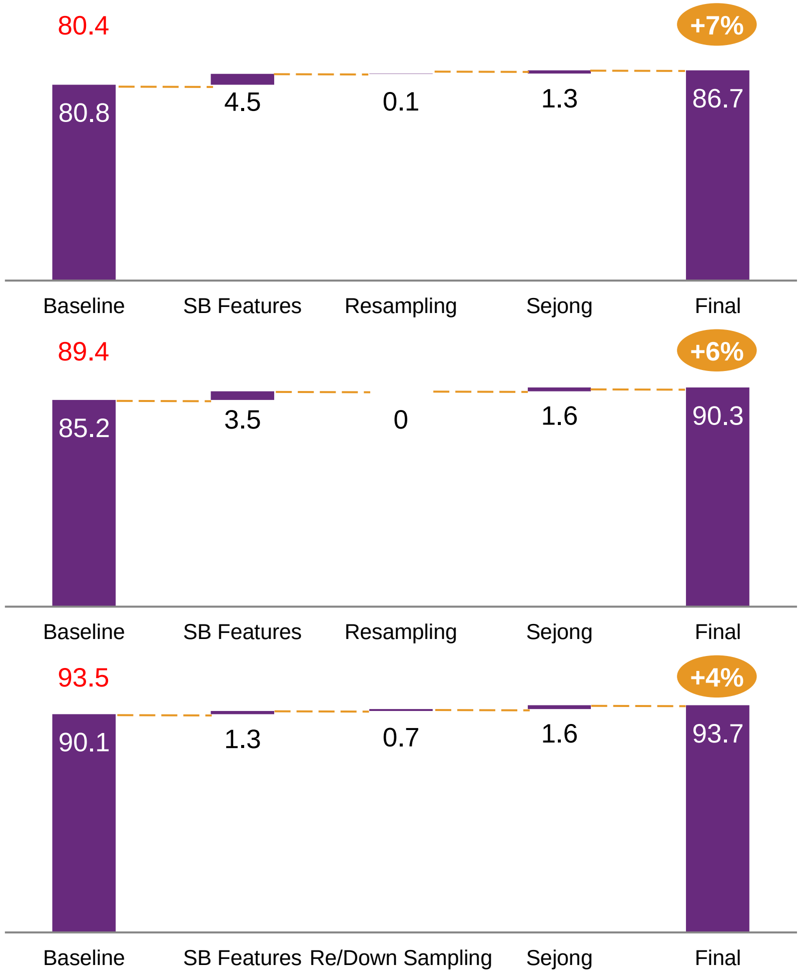
XX Reference performance from journal   XX % improvement from our baseline

Overview of model AUROC

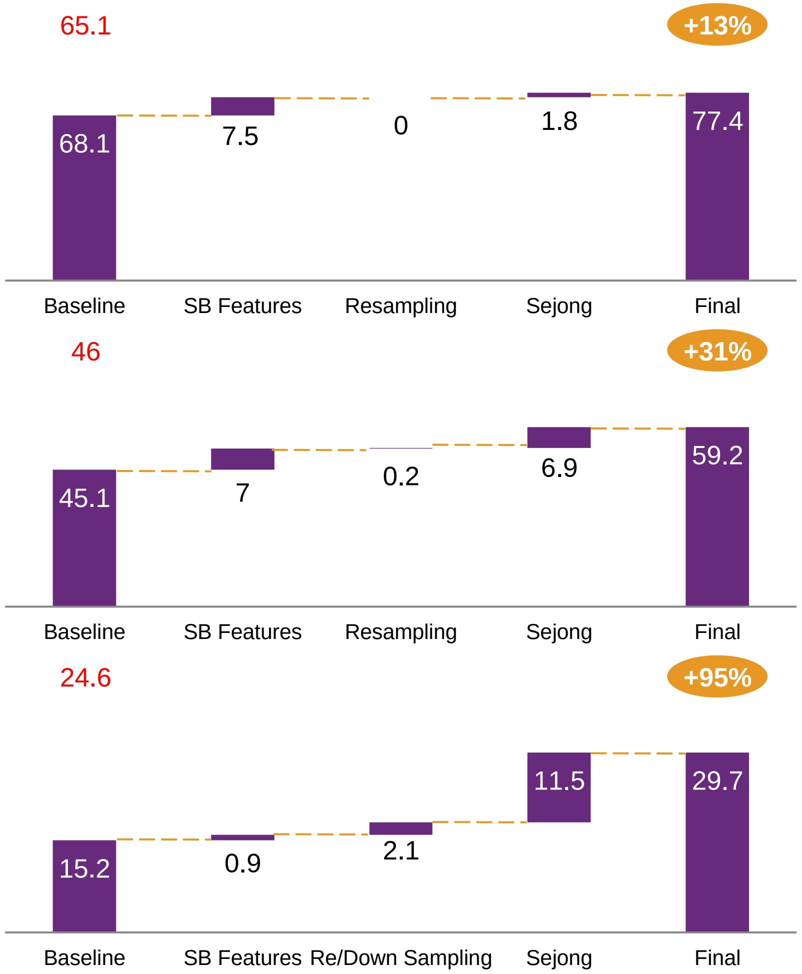
ADMISSION

CRITICAL

DEATH



Overview of model AUPRC



OBSERVATIONS

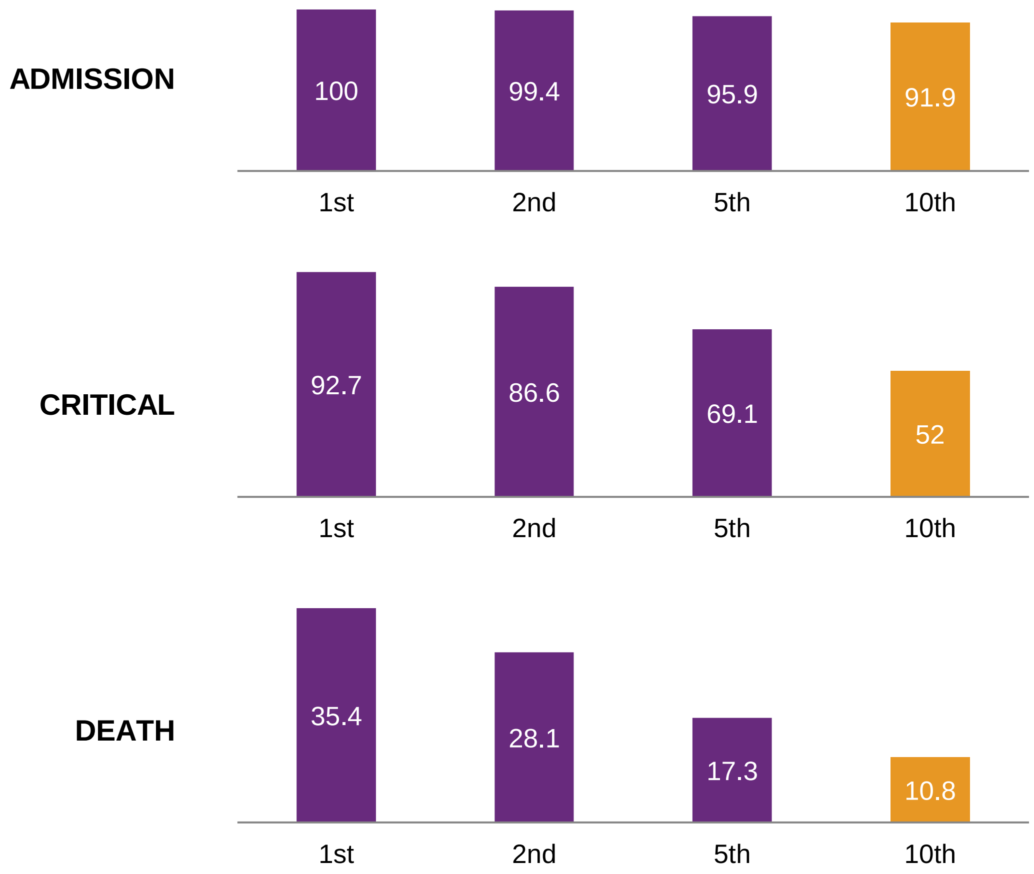
- We are able to achieve better AUROC compared to baseline and similar AUROC compared to the reference journal utilising deep learning
- We see large improvements in AUPRC against both our baseline and reference journal, which suggests the models have **better predictive power for the severe cases**
- Admission and critical cases benefit the most from SB features introduced while deaths benefit most from Sejong data
- The models trained and tested on Sejong perform better than resampled NEDIS data, which suggest there is some fundamental difference between the patients in NEDIS and Sejong beyond age

Iteration 1

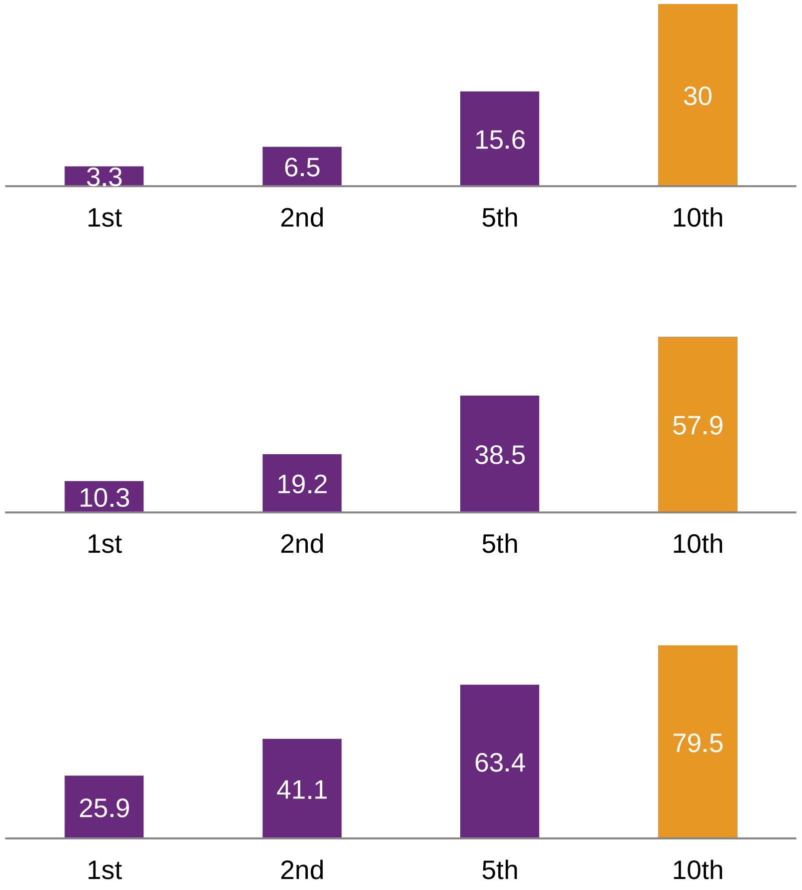
Modelling Outcomes (2/2)

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Cumulative precision of final model at percentile (%)



Cumulative recall of final model at percentile (%)



OBSERVATIONS

- The death model is able to identify ~80% of all death cases by the 10th percentile, which is the most severe and therefore requires the most medical attention
- For critical cases, at the 10th percentile we are able to identify 50% of the patients correctly and ~60% of all critical cases
- For admissions, we are able to be targeted with over 90% of all patients correctly identified at the 10th percentile

Internal Consistency

Model prediction rates from Sejong test at 0.5 probability cut-off vs target prevalence in NEDIS

Death	Critical	Admission	Prevalence in NEDIS (%)	Sejong Predictions (%)
1	1	1	2	1
1	1	0	0	0
1	0	0	0	0
1	0	1	0	1
0	1	1	7	12
0	0	0	66	75
0	1	0	0	0
0	0	1	25	21

OBSERVATIONS

- The 3 model approach provides fairly consistent results with around only 1% invalid outcomes
- However, this does not tell us if using 3 models will be easier to interpret the final outcome and act on it – for that we will be comparing this against a three-class classification model using the new target variable

Iteration 3

New Target Comparison

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Comparison of 3 model performance vs new target

3 model results (overall accuracy = 76%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/8
http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/20

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	896	45	841	1782
CRIT	403	189	146	738
NONE	512	5	5191	5708
	1811	239	6178	8228

ACTUAL	ADM	CRIT	NONE	
ADM	11%	1%	10%	22%
CRIT	5%	2%	2%	9%
NONE	6%	0%	63%	69%
	22%	3%	75%	100%

Favourable for admission (Biased against critical cases)

New target results (overall accuracy = 77%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong/12

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	690	122	970	1782
CRIT	213	313	212	738
NONE	331	45	5332	5708
	1234	480	6514	8228

ACTUAL	ADM	CRIT	NONE	
ADM	8%	1%	12%	22%
CRIT	3%	4%	3%	9%
NONE	4%	1%	65%	69%
	15%	6%	79%	100%

Favourable for critical (Biased against admission cases)

Sejong will need to trade-off between better identifying admission and critical cases when deciding which model to use

Iteration 3

New Target Comparison (Tuned)

Problem Definition | Data Understanding | Data Preparation | Insights & Validation | Modelling & Evaluation

Comparison of 3 model performance vs new target

3 model results (overall accuracy = 76%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/8
http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/20

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	896	45	841	1782
CRIT	403	189	146	738
NONE	512	5	5191	5708
	1811	239	6178	8228

ACTUAL	ADM	CRIT	NONE	
ADM	11%	1%	10%	22%
CRIT	5%	2%	2%	9%
NONE	6%	0%	63%	69%
	22%	3%	75%	100%

Favourable for admission (Biased against critical cases)

New target results (overall accuracy = 76%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong/27

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	881	133	768	1782
CRIT	259	328	151	738
NONE	612	51	5045	5708
	1234	480	6514	8228

ACTUAL	ADM	CRIT	NONE	
ADM	11%	2%	9%	22%
CRIT	3%	4%	2%	9%
NONE	7%	1%	61%	69%
	15%	6%	79%	100%

Favourable for critical (Biased against admission cases)

Tuning the cut-off helps reduce the number of miss classified critical patients in the new target model

Iteration 3

Comparison From Baseline

Evolution of model performance

Baseline model (accuracy = 72%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/25
http://52.141.17.15/#/visualPipeline/Usecase1_ER_Sejong_method/24

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	731	34	1017	1782
CRIT	363	117	258	738
NONE	565	11	5132	5708
	1811	239	6178	8228

3 model results (accuracy = 76%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/8
http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong_CV_pb/20

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	896	45	841	1782
CRIT	403	189	146	738
NONE	512	5	5191	5708
	1811	239	6178	8228

New target results (accuracy = 76%)

http://52.141.17.15/#/visualPipeline/Usecase1_ER_modelling_train-on_Sejong/27

PREDICTED

ACTUAL	ADM	CRIT	NONE	
ADM	881	133	768	1782
CRIT	259	328	151	738
NONE	612	51	5045	5708
	1234	480	6514	8228

180% improvement

We are able to predict 211 more critical cases correctly compared to baseline

Iteration 3

Interpretability of Models

Problem Definition → Data Understanding → Data Preparation → Insights & Validation → Modelling & Evaluation

Sample model predictions

Overlapping probabilities

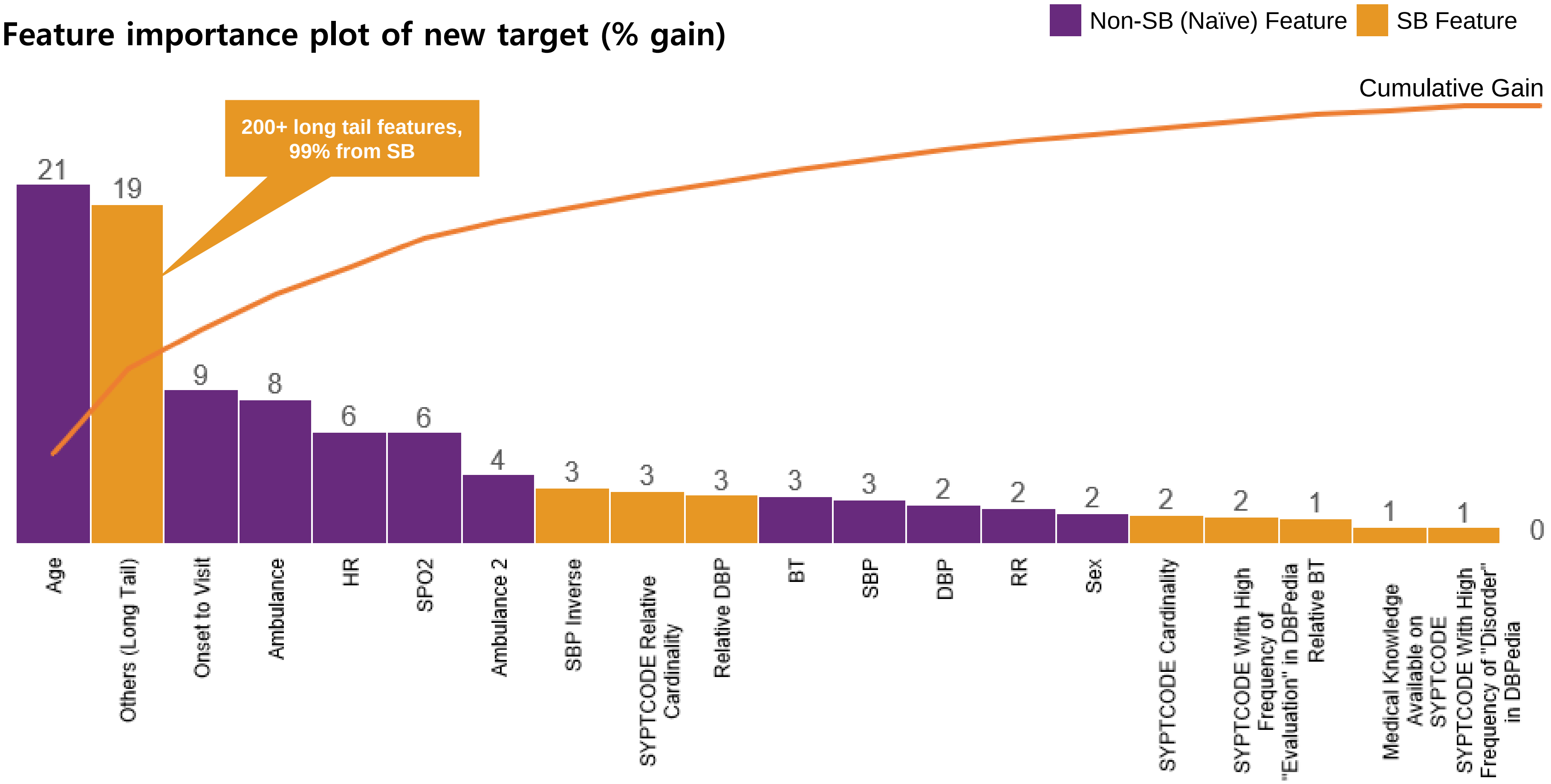
Actual			3 Model Predictions			New target model predictions		
death	crit	adm	death	crit	adm	crit	adm	none
0	0	1	0.2	0.1	1.0	0.9	0.0	0.0
0	0	1	0.2	0.2	0.9	0.8	0.1	0.1
0	1	1	0.1	0.4	0.9	0.4	0.5	0.1
0	0	0	0.2	0.4	0.8	0.4	0.5	0.1
0	0	0	0.1	0.5	0.8	0.3	0.5	0.2
1	1	1	0.6	0.6	1.0	0.2	0.8	0.0
0	1	1	0.2	0.2	0.9	0.5	0.3	0.1

Paralysis by analysis

The new target provides clearer indication for physicians by eliminating ambiguity

Feature Importance

Feature importance plot of new target (% gain)



OBSERVATIONS

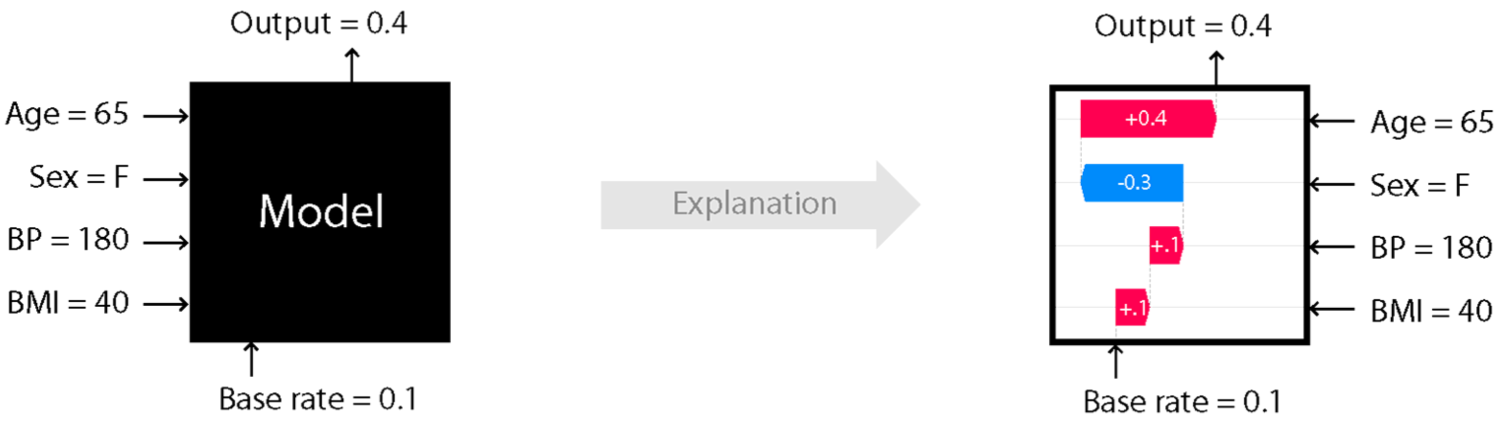
- The model heavily relies on naïve features as observed by the top features emerging from the plot
- While the model places emphasis on naïve features, **8 out of 18 top features contributing to 80% of the model gain** (excluding long tail) **were SB features**
- In totality, **96% (229/244) of all features used** in the model were **generated by the platform**
- These SparkBeyond features contributed to **~32% of total gain in the model**

SparkBeyond features are critical contributors to model predictions

Explaining predictions (1/2)

Overview of SHAP (SHapley Additive exPlanations)

SHAP



What is SHAP?

- A game theoretic approach to explain the output of any machine learning model.
- Connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions

Why should Sejong use this?

- SHAP can help doctors explain why certain predictions were made for a specific patient
- Doctors can use this to augment their own knowledge and KTAS scores to make the final decision
- SHAP is different from regular feature importance, which is usually aggregated for the entire model while SHAP can provide feature importance for individual predictions

Operationalization

Explaining predictions (2/2)

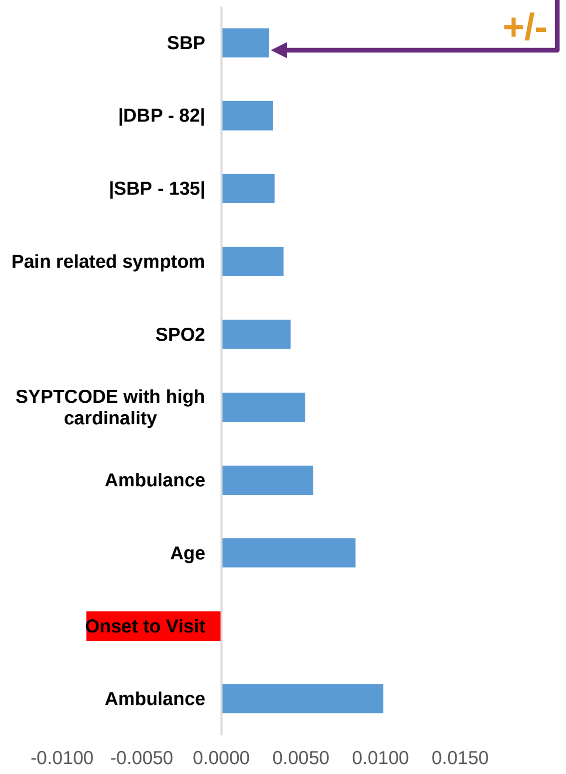
Problem Definition | Data Understanding | Data Preparation | Insights & Validation | **Modelling & Evaluation**

Example of Top 10 SHAP values for binary critical model predictions

Patient 1

Actual: 0 Base Value: 0.91

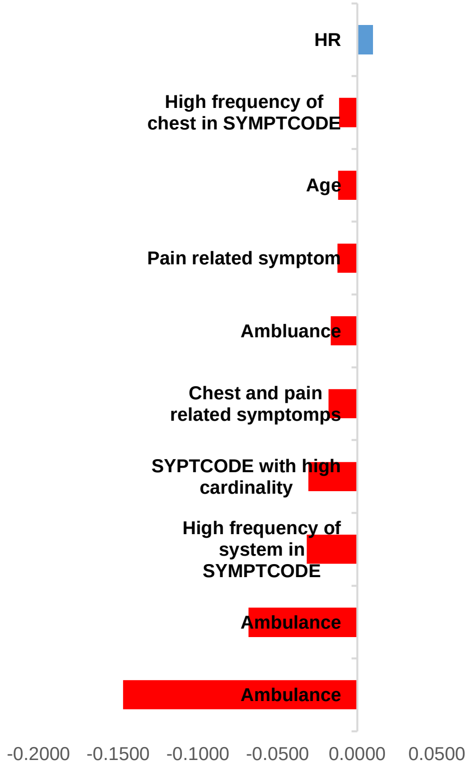
Predicted: 0 Predicted Probability: 0.99



Patient 2

Actual: 1 Base Value: 0.91

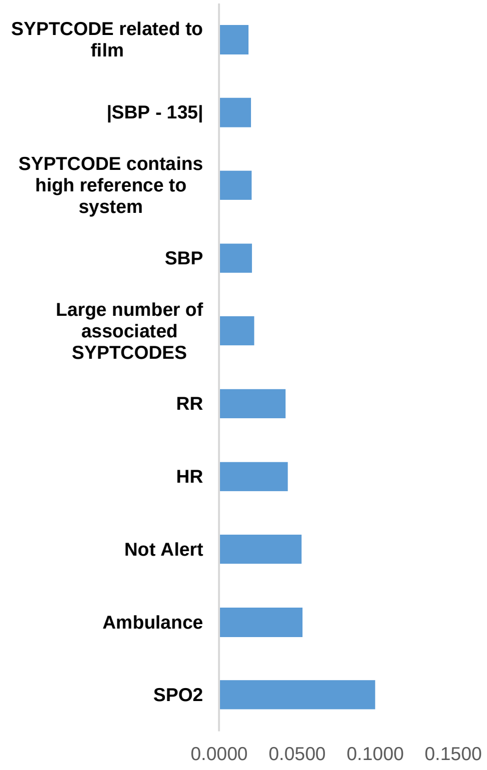
Predicted: 0 Predicted Probability: 0.57



Patient 3

Actual: 1 Base Value: 0.09

Predicted: 1 Predicted Probability: 0.64



OBSERVATIONS

- SHAP values can help augment physician judgement when deciding on the right course of treatment – one can either affirm or deny the model predictions by looking at important variables
- SHAP values become important when Sejong operationalizes the model as no physician would solely rely on machine learning nor **rely on predictions without explanations**

Suggestions for future work

Suggested ideas

Data

- Get additional symptom codes from NEDIS
- Time series of triage tests
- Patient history
- Augment with Medical subject headers
- Augment with google search trends

Insights

- Multi-hopping UMLS topics within DBpedia for deeper topical search
- Multi-hopping patient history
- Multi-hopping data from other hospitals

Modelling

- Feature selection
- Removing and validating duplicate features
- Hyperparameter tuning of new target model
- Replicating exact deep learning parameters used in the model with SB features

NEXT STEPS

- We will work with you to identify how we can **operationalise this model** at Sejong hospital

There are many more permutations of experiments to be run!

Platform

Develop a physician friendly dashboard for explainable model predictions

Overview of Dashboard

NEXT STEPS

- We will work with you to identify how we can develop a user-friendly dashboard capable of explainable AI

We need to involve ER physicians to make this useful and attractive to them