Healthcare in Ireland: Reduce hospital wait time and prevent stroke & diabetes

Data Collection Source and attributes Classification Diabetes Health Indicators Dataset (Teboul, 2021) Data Attributes- Stroke Veggies HighBP HvyAlcoholConsump HeartDiseaseor HighChol AnyHealthcare Attack CholCheck **Regression & Clustering** PhysActivity NoDocbcCost BMI Fruits GenHlth Smoker PhysHlth DiffWalk MentHlth Sex Education &Income IPDC Waiting List By Group Hospital 2018-2020 (ehealth Ireland, 2014-2020). Data Attributes- Archive Date Speciality Stroke Prediction Dataset (Fedesoriano, 2021) Group Case type Data Attributes- Hospital HIPE Adult/Child work_type id Hospital Age Categorisation Residence_type Speciality HIPE gender Time Bands avg_glucose_level age Count/Total bmi hypertension smoking_status heart_disease stroke

Data Pre-Processing

ever_married

- Hospital Name Correction Hospital name have changed over year. "Letterkenny General Hospital" changed to "Letterkenny University Hospital", "Lourdes Orthopaedic Hospital Kilcreene" changed to "Kilcreene Regional Orthopaedic Hospital".
- Filling Null value/Missing Data For missing BMI values, we have used the mean BMI of the dataset to impute these missing entries. For patient wait times missing month data, we have assigned null values.
- **Dropping Duplicates** In the dataset concerning Diabetes and Stroke, we identified duplicate records and subsequently removed these duplicates.
- Data Scaling & Normalisation The Z-score normalization technique was used on the 'Total' attribute to standardize it. Additionally, we used the Standard Scaler method on other attributes to scale them to a mean of zero and a standard deviation of one.
- Fix Coloum Name Identify errors in data such as attribute name changed from "count" to "total" in Waiting List data.

Algorithms & Models

REGRESSION

- Linear regression
- Naïve model
- Simple average forecast
- (2/4/7/10) point trailing moving average
- Simple Exponential Smoothing

CLUSTERING

- K-means algorithm
- DB-scan

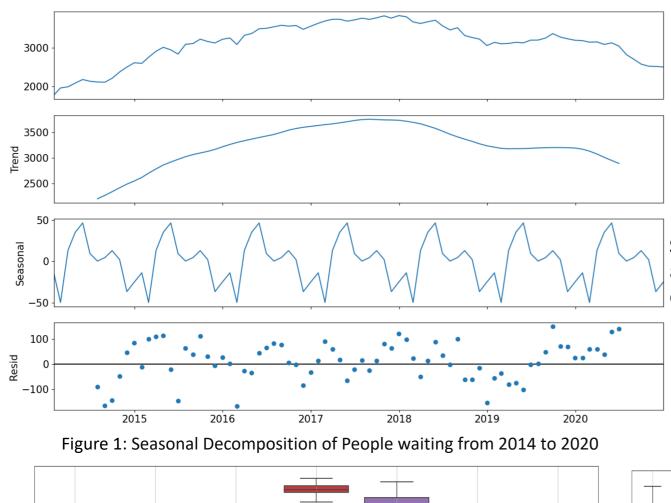
CLASSIFICATION

- K-nearest neighbors
- Random forest classifier

Results

Forecasting number of patients on waitlist- Regression

For regression, we focused on University Hospital Galway, examining inpatient cases across all medical specialties.



OBSERVED DATA: shows the actual recorded values of the dataset, with possible underlying trends and seasonal effects present, ranging from 2014 to 2020

TREND: illustrates a smooth, long-term increase in the total number of patient waiting, peaking around 2018, suggesting a general upward movement over the years.

SEASONAL FLUCTUATIONS: component indicates a repeating pattern of patient waiting time that occurs regularly within each year.

RESIDUAL COMPONENT: displays the irregularities or 'noise' left in the patient waiting data after the trend and seasonal components have been removed, highlighting deviations not explained by the model.

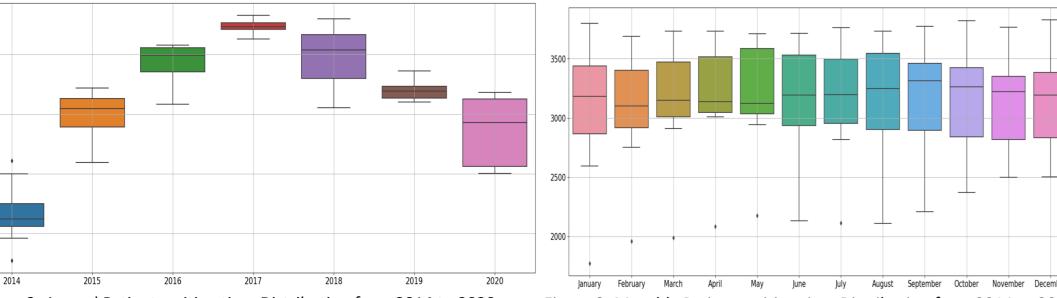


Figure 2: Annual Patient waiting time Distribution from 2014 to 2020 Figure 3: Monthly Patient waiting time Distribution from 2014 to 2020

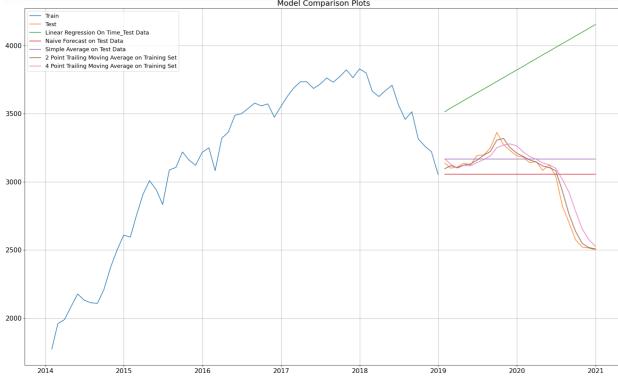


Figure 4 compares different forecasting models we

Green Line indicates Linear Regression with much higher RMSE value of 911 approx.

Two horizontal line red and purple are results of Naïve forecasting technique and Simple average forecasting technique producing RSME of 260.5 and 294 (approx.) respectively.

Brown and Pink line graph are 2 point trailing moving average and 4 point trailing moving average technique with RMSE value of 37.8 and 89.9 (approx.) respectively.

Figure 4: Comparative Forecasting Models for Patient Waiting Time in 2021

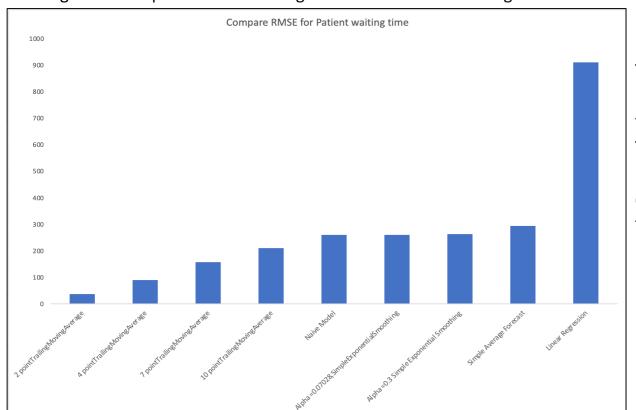
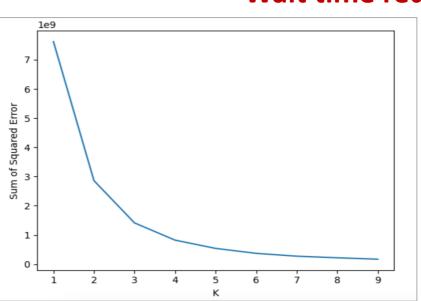
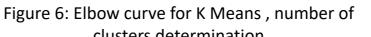


Figure 5: RMSE Comparison of Forecasting Models for Patient Waiting Times

The 2-point trailing moving average has the lowest RMSE, indicating it provided the most accurate forecast for total patient waiting dataset. The significant variation in performance between the linear regression showing a much higher RMSE compared to the others, indicating less accuracy in the context of this forecast.

Wait time reduction - Clustering





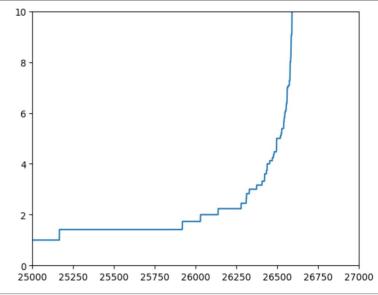


Figure 7: Curve to find out the optimum epsilon due for Dhiscan clustering algorithm

clusters determination		value for Db scan clustering algorith			
		Number of clusters	Silhouette score	Min _ Silhouette width	Max_Silhouette Width
K-Means Clustering	With Z normalised	k=3	0.3481	-0.114	0.54
	Not normalised	k=3	0.8274	-0.0007	0.92
DB_Scan (Scaled data)					
	eps=4, min_samples=10	2	0.728	-0.695	0.948
	eps=4, min_samples =1000	2	0.9375	-0.768	0.967
	eps=2.1, min_samples =10	3	-0.2004	-0.768	0.96
	eps=2.1, min_samples =1000	3	-0.168	0.967	-0.76

Figure 8: Results of Clustering Algorithms

- Silhouette scores, minimum and Silhouette width (Figure 8) are considered to evaluate the models.
- Db_scan with epsilon value of 4 and minimum samples in a cluster to be 1000, has the highest silhouette score of 0.93, however the minimum silhouette width is -0.768 and there are only 2 clusters formed.
- We decide to choose k-means as it has 3 clusters and 0.82 silhouette score with least minimum silhouette width of -0.0007, which indicates good separation between the clusters.

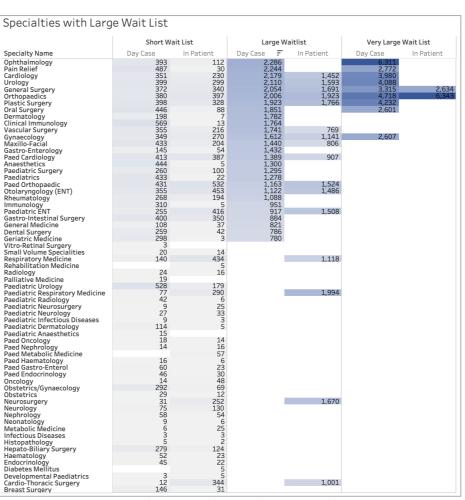


Figure 9:Specialties with very large waitlist as clustered

Figure 10: Hospitals with very large waitlist as

clustered

2,135 2,246 1,874 1,974

2,415

1,154 1,387

1,031

886 2,303

849 1,426

Hospitals with large wait list

Royal Victoria Eye and Ear Hosp University Hospital Waterford

Nenagh Hospital

St. Vincent's University

Wexford General Hospita

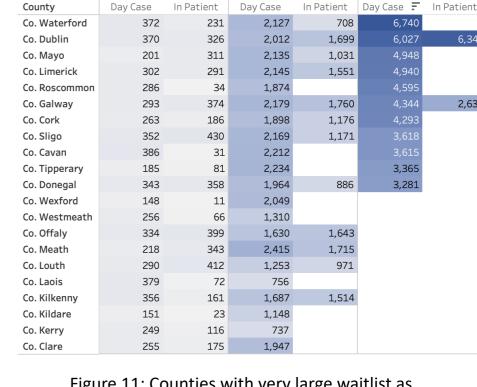
St. Luke's General Hospital Kilk

South Tipperary General Hos

Our Lady of Lourdes Hospital Droghed

Our Lady's Hospital Navar

Louth County Hospital



Large Waitlist

Figure 11: Counties with very large waitlist as clustered

Recommendations

•Increase bed capacity:

Counties with Large Wait List

- National Orthopaedic Hospital Cappagh (81 -2014,79 -
- University Hospital Galway (644 beds-2014, 668 2020) Share best practices

•St. James Hospital in Dublin has "Short Wait List" despite being a big hospital and constant bed count over the years. •Increase diagnostic & laboratory capabilities - for increased diabetes and stoke screening

•Specialists Recruitment and Retention:

•Hire more orthopaedic surgeons, ophthalmologists and surgeons to counter demand for these specialties. •Allocate resources and prioritize interventions in reproductive health care.

•Allocate resources to support training and education of healthcare professionals in specialties with high demand.

•Construct new outpatient clinics to counter high no. of day cases in Waterford and Dublin.

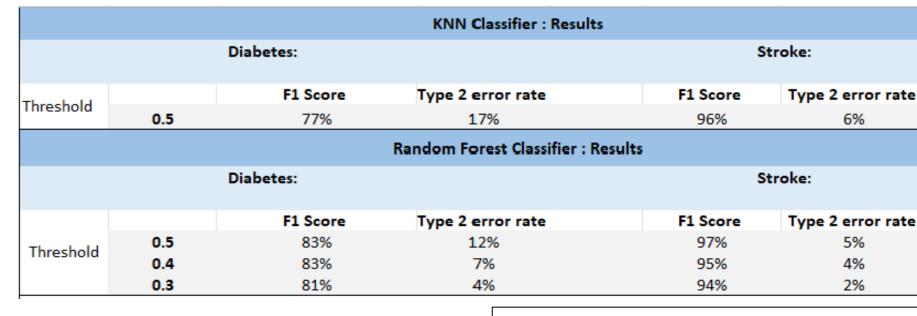
•Invest in Tele-medicine and community-based care for high day case waitlist (Waterford & Dublin)

High Risk Patient identification- Classification

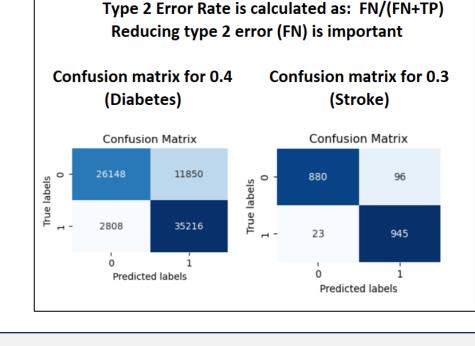
We focused on predicting the risk of diabetes and stroke as the numbers for these healthcare issues are concerning.

Very Large Wait List

We tried K-nearest neighbour and Random Forest Classifiers.



- Random forest classifier seems to perform better with better F1 scores and Type 2 error rates.
- Our metric under measurement was Type 2 error as we wanted to reduce the number of False Negatives in the healthcare industry.
- Further, we tried reducing the Type 2 error by reducing the classification criteria threshold.
- The best threshold for diabetes was found to be 0.4 and for stroke, it was 0.3.



References:

ehealth Ireland (2014) IPDC Waiting List By Group Hospital 2014 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/fd1d1a0e-818d-4a5f-9a54-de8f176cccdd (Access ehealth Ireland (2016) IPDC Waiting List By Group Hospital 2016 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/cab14262-039f-49d4-a57b-68b4b98a957d ehealth Ireland (2017) IPDC Waiting List By Group Hospital 2017 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/0319599c-6fd0-42fb-a27d-762e1cada604 (Accessed in the control of the cont ehealth Ireland (2018) IPDC Waiting List By Group Hospital 2018 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/35f3488e-8fd3-45af-8456-

ehealth Ireland (2019) IPDC Waiting List By Group Hospital 2019 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/8a28fc2f-717e-4975-9771-2f129abad36a (Accessed

ehealth Ireland (2020) IPDC Waiting List By Group Hospital 2020 - HSE Open Data (online). Available at: https://data.ehealthireland.ie/dataset/ipdc-waiting-list-by-group-hospital/resource/88d51b97-aecd-48ae-9ca9-c28d28d65f4e (Accessed and Accessed and

FEDESORIANO (2021) 'Stroke Prediction Dataset (online)'. Available at: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset (Accessed 20 April 2024). [EBOUL (2021) 'Diabetes Health Indicators Dataset (online)'. Available at: https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset (Accessed 20 April 2024).