

Analysis of Sepsis

Rachel H-Blume

March 2023

Executive Summary:

This study took data analytics and merged it with machine learning as they challenged groups to develop algorithms that would diagnosis sepsis quicker for more timely intervention and better patient outcomes. They evaluated the algorithms for timeliness and then took away points when there was a delay intervention or false positives.

These were openly sourced entries from three different geographic locations. I was interested in this because I am a clinician that takes care of sceptic patients and I am now learning data analytics.

Analysis of Sepsis

Rachel H-Blume
March 2023

Data		The data source was Kaggle. The data set I linked here is called "Prediction of Sepsis" and was last updated a year ago. Salik Hussaini is the owner of the set. These values were reported hourly for the patients.
Data Source		This is public data downloaded from Kaggle.com
Data Link		https://www.kaggle.com/datasets/salikhussaini49/prediction-of-sepsis VERSIONS/2?resource=download
Data Collection		3 Separate Hospitals, over 60,000 patients and 40 variables per patient, with 3 specific Sepsis Criteria set, over 853 submitted data and they used a special analysis to compare it
Data Profile		41 rows of time series data was calculated for each patient. This was broken into category such as demographic, laboratory, numeric clinical vitals, and patient outcome designation. 1552210 Rows of data in 44 Columns Cleaned to 32 Rows

Analysis of Sepsis

Rachel H-Blume

March 2023

HR	Heart Rate (bpm)	Discrete, Quantitative	Time Variant
02 Sat	Pulse Oximeter (%)	Discrete Quantitative	Time Variant
Temp	Degree (c)	Continuous Quantitative	Time Variant
SBP	Systolic BP (mm Hg)	Discrete Quantitative	Time Variant
MAP	Mean Arterial Pressure (mm Hg)	Discrete Quantitative	Time Variant
DBP	Diastoloc BP (mm Hg)	Discrete Quantitative	Time Variant
RR	Respiration Rate (RR) breaths per minute	Discrete Quantitative	Time Variant
BE	Base excess bicarbonate (mmol/L)	Continuous Quantitative	Time Variant
HCO3	Bicarbonate (mmol/L)`	Continuous Quantitiative	Time Variant

Analysis of Sepsis

Rachel H-Blume

March 2023

FIO2 (%)	Fraction of inspired Oxygen (O2%)	Discrete Quantitative	Time Variant
pH	pH	Continuous Quantitative	Time Variant
PaCO2	Partial pressure of carbon dioxide from arterial blood (mm Hg)	Continuous Quantitative	Time Variant
SaO2	Oxygen Saturation from arterial blood (%)	Continuous Quantitative	Time Variant

Lactate	Lactic Acid (mg/dL)	Continuous Quantitative	Time Variant
Bilirubin total	Total Bilirubin (mg/dL)	Continuous Quantitative	Time Variant
Hct	Hematocrit (%)	Continuous Quantitative	Time Variant
Hgb	Hemoglobin (g/dL)	Continuous Quantitative	Time Variant
PTT	Partial thromboplastin time (seconds)	Continuous Quantitative	Time Variant
WBC	Leukocyte count (count/L)	Continuous Quantitative	Time Variant

Analysis of Sepsis

Rachel H-Blume

March 2023

Fibrinogen	Fibrinogen concentration (mg/dL)	Continuous Quantitative	Time Variant
Platelets	Platelet Count (count/mL)	Continuous Quantitative	Time Variant
Magnesium	Magnesium (mmol/dL)	Continuous Quantitative	Time Variant
Age	Years	Discrete Quantitative	Time Variant
Gender	0- Female 1=Male	Discrete Qualitative	Time Invariant
Unit 1 Administrative Identifier for ICU (MICU)	0= False 1= True	Discrete Qualitative	Time Variant
Unit 2 Administrative Identifier for ICU (SICU)	0=False 1=True	Discrete Qualitative	Time Variant
HospAdmin Time	Time between hospital admission and ICU admission (hours)	Continuous Quantitative	Time Variant
ICULOS ICU	Hours since ICU admission	Continuous Quantitative	Time Variant
SepsisLabel for Septic Patients	Sepsis Label is: 1 if $t \geq t_{sepsis} - 6$ 0 if $t < t_{sepsis} - 6$ Non-septic Sepsis Label is 0	Discrete Qualitative	Time Variant

Analysis of Sepsis

Rachel H-Blume

March 2023

Data Cleaning:

Dropped the following columns as i will not be using them for my analysis:

- End Tidal CO₂
- Alkalinephos,
- Calcium, Chloride
- Creatine
- Bilirubin-direct
- Glucose

- *AST
- *BUN
- *Phosphate
- *Potassium
- *Troponin

NO Duplicates Noted

Took the Mean Value and replaced the following NaNs

- a. HR
 - O₂Sat
 - SBP,MAP,DBP
 - Resp
 - Base Excess
 - HcO₃
 - FIO₂
 - pH
 - Platelets

- * PaCO₂
- *SaO₂
- *Lactate
- *Hct
- *Hgb
- *PTT
- *WBC
- * Fibrinogen

I purposefully left the Unit Nans because they do not change any calculations

Analysis of Sepsis

Rachel H-Blume

March 2023

Data Limitations/Ethics:

This data had many missing values reported above that I used the Mean for. The data for each patient was reported hourly so it probably did not change a huge amount so average should be fine.

The investigators did multiple locations so that is nice as it will include geography later. They also looked at programming and plagiarism. Overall this is a robust set with ethical guidelines followed as patient identifiers are gone and it was well cited.

Investigative Questions

- 1: Does FiO₂ in Sepsis vary by gender?
2. Does Sepsis frequency vary by gender?
3. Does pH really indicate Sepsis ?