

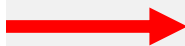


# Final Project : Fetal Health Classification

Oleh : Subhany



# Latar Belakang Pemilihan Judul



Tujuan No. 3:  
Kehidupan  
Sehat dan  
Sejahtera  
Tahun 2030

Tujuan SDGs Target  
3.1 pada tahun 2030  
adalah mengurangi  
rasio angka kematian  
ibu hingga kurang dari  
70 per 100.000  
kelahiran hidup

## Realita Indonesia

Nilai AKI Indonesia adalah sebesar 189 per 100.000 kelahiran hidup, sementara Provinsi Papua mempunyai AKI tertinggi (565) dan Provinsi DKI Jakarta mempunyai AKI terendah (48)

## Target Indonesia

Menurunkan rasio  
angka kematian ibu  
hingga kurang dari 70  
per 100.000 kelahiran  
hidup pada tahun 2030



AKI = Angka Kematian Ibu

Sumber: Berita Resmi Statistik No.  
09/01/Th. XXVI, 30 Januari 2023

# Kendala Perolehan Data

1. Tidak ada data Angka Kematian Ibu dan Anak di Kaggle.com
  2. Diperoleh data BPS:
    - a. Angka Kematian Neonatal menurut Daerah Tempat Tinggal
    - b. Angka Kematian Neonatal menurut Kuantil Kekayaan
    - c. Angka Kematian Neonatal menurut Pendidikan Ibu
    - d. Angka Kematian Neonatal menurut Provinsi
    - e. Angka Kematian Neonatal menurut Usia Ibu saat Melahirkan
- Namun kelima data ini tidak dapat digabung karena mempunyai row yang berbeda-beda



Menggunakan Dataset Fetal\_Health.csv from Kaggle.com (well documented and Gold)

# Dataset & Project Overview

Data set “Fetal\_Health” pada Kaggle adalah upaya untuk mengidentifikasi dan menganalisis mengenai faktor-faktor yang mempengaruhi kesehatan janin berdasarkan detak jantung janin, intervensi yang diperlukan dan upaya peningkatan kesehatan janin selama masa kehamilan.

Pengukuran kesehatan janin berdasarkan detak jantung dilakukan dengan menggunakan alat Cardiotocography (CTG) yang berfungsi sebagai alat monitor kesehatan detak jantung janin selama masa kehamilan..

Hasil Interpretasi CTG membantu dalam menentukan apakah kehamilan berisiko tinggi atau rendah. Hasil CTG yang tidak normal dapat mengindikasikan perlunya observasi lebih lanjut terhadap kesehatan ibu dan anak serta intervensi yang perlu dilakukan.

Tujuan:

- ☐ Mengidentifikasi faktor-faktor yang mempengaruhi kesehatan janin.
- ☐ Menentukan lebih dini / awal terkait intervensi yang diperlukan dalam meningkatkan kesehatan janin selama masa kehamilan

# Tahapan Pemodelan



## Problem



1. **Identifikasi kesehatan janin berdasarkan fitur-fitur** yang terdapat dalam dataset.
2. Mengidentifikasi **faktor-faktor yang berkaitan dengan kesehatan janin.**



1. **Meningkatkan pemahaman** tentang faktor-faktor yang mempengaruhi kesehatan janin.
2. **Meningkatkan kemampuan** dalam memprediksi atau mendeteksi kondisi kesehatan janin.



## Goals

## Objective



1. **Menganalisis distribusi fitur-fitur** dalam dataset untuk memahami karakteristik data.
2. **Melakukan eksplorasi data** untuk mengidentifikasi hubungan antara fitur-fitur dan kesehatan janin.

1. **Membantu mengukur efektivitas** penggunaan dataset dalam perawatan kesehatan janin
2. **Memahami dampak dari implementasi model prediktif** dalam perawatan kesehatan janin dan membuat keputusan yang lebih baik untuk meningkatkan kualitas perawatan pasien



## Business Metrics



# Data Understanding & Pre-Processing



# Dataset Information

- 2126 rows
- 22 columns
- 13 duplicate rows (dropped)
- No Missing Values
- All columns are numeric
- 1 target
- 8 features



## Column Name & Explanations

**Baseline Value:** adalah nilai-nilai dasar atau rata-rata dari parameter yang diukur, mengacu pada data seperti tekanan darah atau detak jantung yang normal selama kehamilan.

**Acceleration:** Ini mengacu pada percepatan detak jantung janin dalam respons terhadap rangsangan.

**Fetal Movement:** Merupakan ukuran dari gerakan janin selama periode waktu tertentu.

**Uterine Contractions:** Ini mengacu pada kontraksi rahim selama kehamilan.

**Light Decelerations, Severe Decelerations, Prolonged Decelerations:** Semua ini mengacu pada perubahan dalam detak jantung janin yang menunjukkan berbagai tingkat keparahan.

**Abnormal Short Term Variability dan Mean Value Short Term Variability:** Variabilitas ini mengacu pada variasi dalam detak jantung janin yang diukur dalam periode waktu yang singkat.

**Percentage of Time with Abnormal Long Term Variability dan Mean Value of Long Term Variability:** Ini adalah ukuran variabilitas detak jantung janin dalam jangka waktu yang lebih panjang.

**Histogram Width, Min, Max, Number of Peaks, Number of Zeroes, Mode, Mean, Median, Variance, Tendency:** Semua ini adalah statistik yang terkait dengan distribusi detak jantung janin.

**Fetal Health:** Ini mengacu pada penilaian umum tentang kesehatan janin, yang dapat mencakup berbagai faktor.

# Data Pre-Processing

Meliputi langkah-langkah seperti :

- **Penghapusan Null / NaN Values**

Pada dataset ini, tidak terdapat Null / NaN Values.

- **Penghapusan Duplicate Values**

Pada dataset ini terdapat 13 Duplicate Values, Row Sample awal

- dataset kita adalah 2126 kemudian karena penghapusan duplicate rows, maka row kita berubah menjadi 2113.

- **Pemisahan data menjadi Training dan Test Set.**

- **Outliers sangat kecil sehingga tidak mengganggu pemodelan**

- **Menetapkan feature dan target dari model**

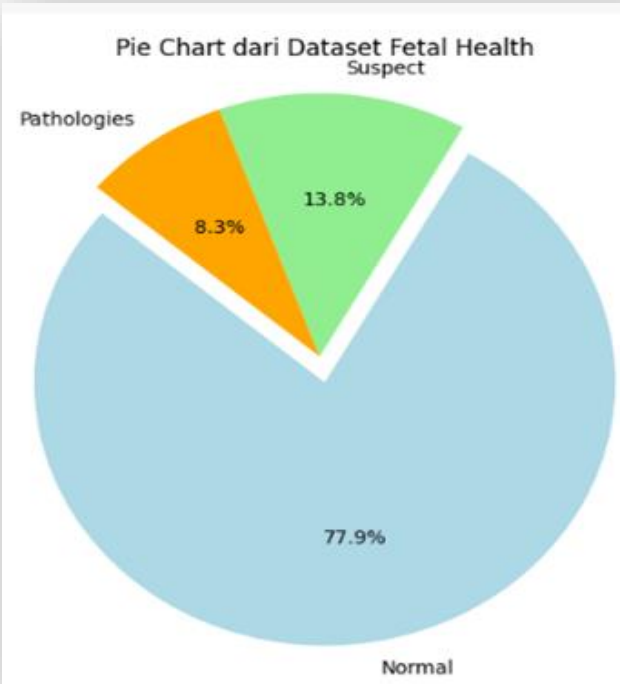






# Exploratory Data Analysis

# Target Column Analysis



77,9%

Normal

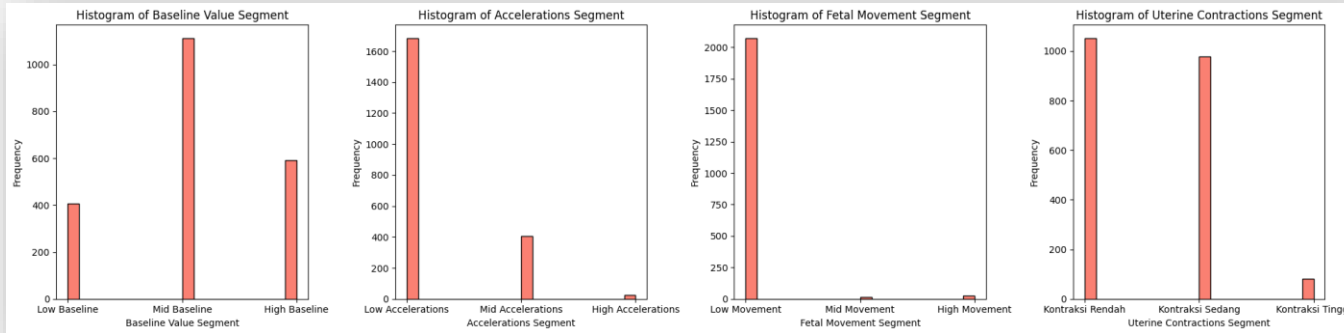
13,8%

Suspect

8,3%

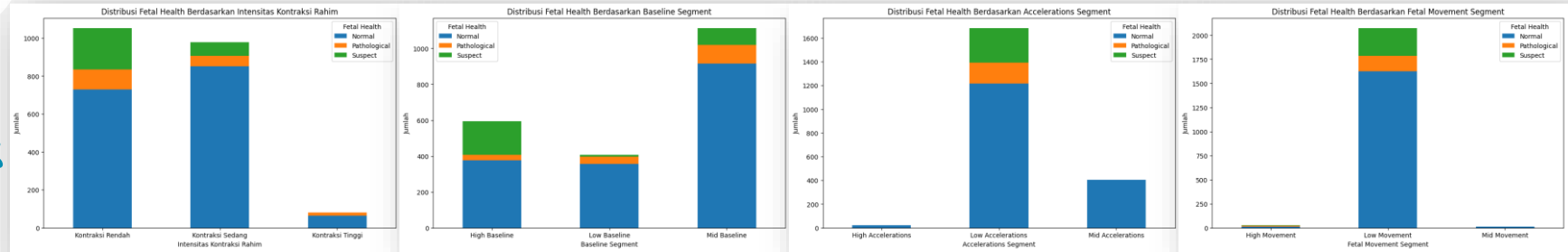
Pathologies

# Segmented Column Analysis



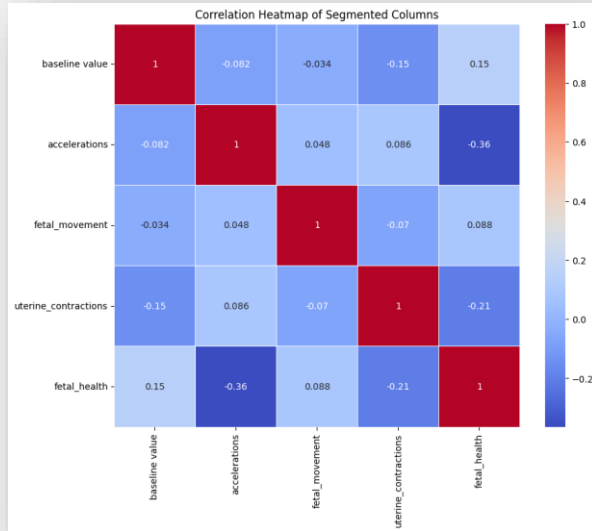
Pada proses ini, kita melakukan Analisa value-value pada kolom yang telah kita segmentasi menggunakan Histogram untuk mengetahui sebaran data / sample pada kolom kita yang telah kita segmentasi tadi.

# Segmented Column Berdasarkan Fetal Health Analysis



Pada proses ini, kita melakukan Analisa value-value pada kolom yang telah kita segmentasi dan kita perbandingkan dengan value pada kolom fetal-health untuk mengetahui apakah ada suatu pola / informasi yang kita dapatkan dari Analisa tersebut. Pada proses ini kita menggunakan '**Bar Chart**' untuk menganalisa kolom-kolom kita

# Analisa Korelasi Kolom Segmentasi Terhadap Target Column



Disini kita akan melakukan Analisa menggunakan '**Heat Map**' untuk mengetahui **korelasi** dari kolom-kolom yang sudah kita segmentasi terhadap target column yaitu '**Fetal Health**'



# Modelling



# Ridge Regression dan Lasso Regression



- Ridge regression dan Lasso regression, memberikan hasil yang sangat serupa dalam hal koefisien fitur untuk dataset ini.
- Keduanya menunjukkan bahwa fitur gerakan janin, variabilitas jangka pendek yang tidak normal, dan persentase waktu dengan variabilitas jangka panjang yang tidak normal memiliki pengaruh yang signifikan terhadap prediksi kesehatan janin, sedangkan fitur-fitur histogram memiliki pengaruh yang lebih kecil atau kurang signifikan.
- Dengan demikian, dalam kasus ini, tidak ada perbedaan besar dalam penggunaan salah satu metode di atas yang dapat dianggap lebih unggul dalam memprediksi kesehatan janin.





# Alogaritma

Final Project ini menggunakan algoritma **Machine Learning** seperti:

- **K-NN Classification**
- **Random Forest**
- **Support Vector Machine**

untuk menganalisis data CTG dengan mudah dan mengevaluasi hasil kesehatan janin dan ibu sebagai **Normal, Suspect, dan Patologis**.







# K-NN Classification

```
[ ] # Prediction and accuracy  
y_pred = knn.predict(X_test)  
accuracy_score(y_test, y_pred) * 100
```

```
93.85342789598108
```

- Dengan nilai akurasi sebesar 93.85%, ini berarti model K-NN memiliki kinerja yang baik dalam melakukan prediksi kelas terhadap data uji.
- Ini menunjukkan bahwa sekitar 93.85% dari label kelas pada data uji diprediksi dengan benar oleh model.
- Sebagai catatan, pemilihan nilai K yang optimal dalam K-NN dapat memengaruhi kinerja model; nilai K yang lebih tinggi cenderung menghasilkan model yang lebih sederhana,





# Evaluation Matriks K-NN

## Evaluation Metrics

```
# KNN Classification report (precision, recall, f1-score, support, accuracy)  
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
1.0	0.95	0.97	0.96	341
2.0	0.81	0.70	0.75	54
3.0	0.96	0.96	0.96	28
accuracy			0.94	423
macro avg	0.91	0.88	0.89	423
weighted avg	0.94	0.94	0.94	423

## Analisis dan Kesimpulan:

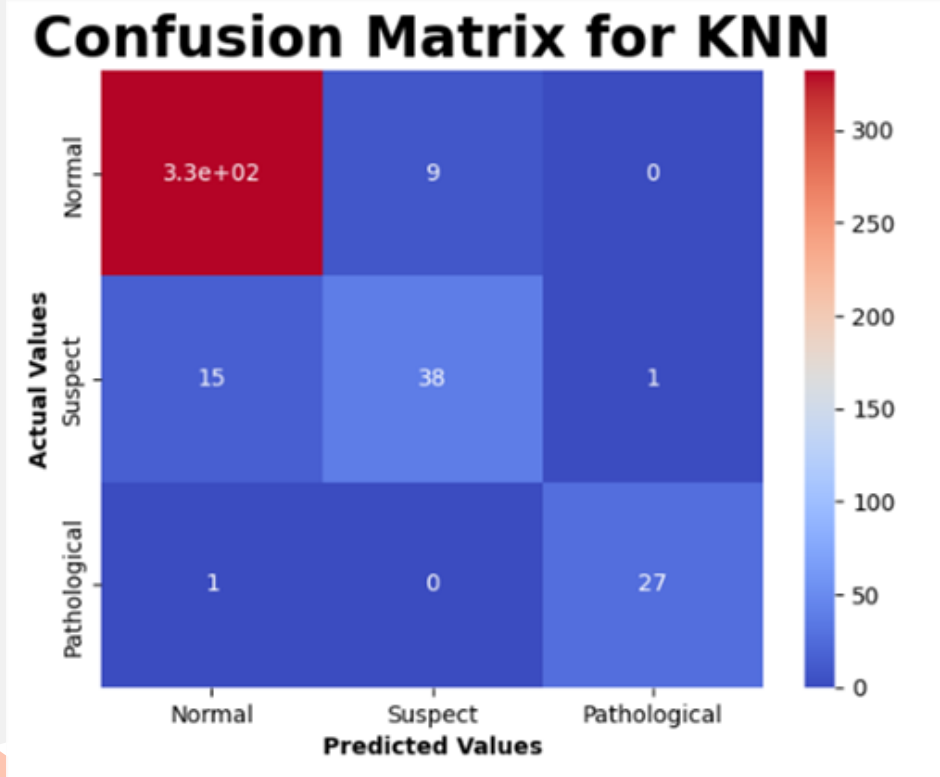
Model memiliki kinerja yang baik dalam mengidentifikasi kelas 1.0 (normal), dengan precision dan recall yang tinggi.

Kelas 2.0 (Suspect) memiliki recall yang lebih rendah, menunjukkan bahwa model memiliki kesulitan dalam mengidentifikasi contoh positif kelas 2.0.

Meskipun kinerja untuk kelas 3 (Patologis) cukup baik, namun terdapat sedikit penurunan dibandingkan dengan kelas 1.0.

Secara keseluruhan, model memiliki akurasi yang baik (90%), tetapi perlu diperhatikan keseimbangan antara precision dan recall untuk setiap kelas agar hasil prediksi menjadi lebih andal.





**\*\*Kesimpulan:\*\***

Bahwa terdapat 300 lebih ibu hamil dengan detak jantung janin normal, dan 38 ibu hamil dengan detak jantung janin dalam kondisi suspect dan 27 ibu hamil dengan kondisi detak jantung janin patalogis.

9 ibu hamil terdeteksi detak janin suspect namun realnya adalah normal dan 15 ibu hamil terdeteksi detak janin normal namun realnya suspect, dan 1 orang ibu hamil terdeteksi normal namun realnya patalog serta 1 orang ibu hami terdeteksi detak jantung janin patalogi namun sesungguhnya adalah suspect





# Random Forest

```
[61] # Prediction dan akurasi  
y_pred_rf = rf_classifier.predict(X_test)  
accuracy_score(y_test, y_pred_rf) * 100
```

```
96.92671394799054
```

Model Random Forest yang diuji pada dataset tersebut memiliki akurasi sebesar 96.93%. Akurasi yang tinggi ini menunjukkan bahwa model memiliki kinerja yang baik dalam memprediksi kelas pada data uji.





# Evaluation Matriks Random Forest

```
# Random Forest Classification report (precision, recall, f1-score, support, accuracy)
print(classification_report(y_test, y_pred_rf))
```

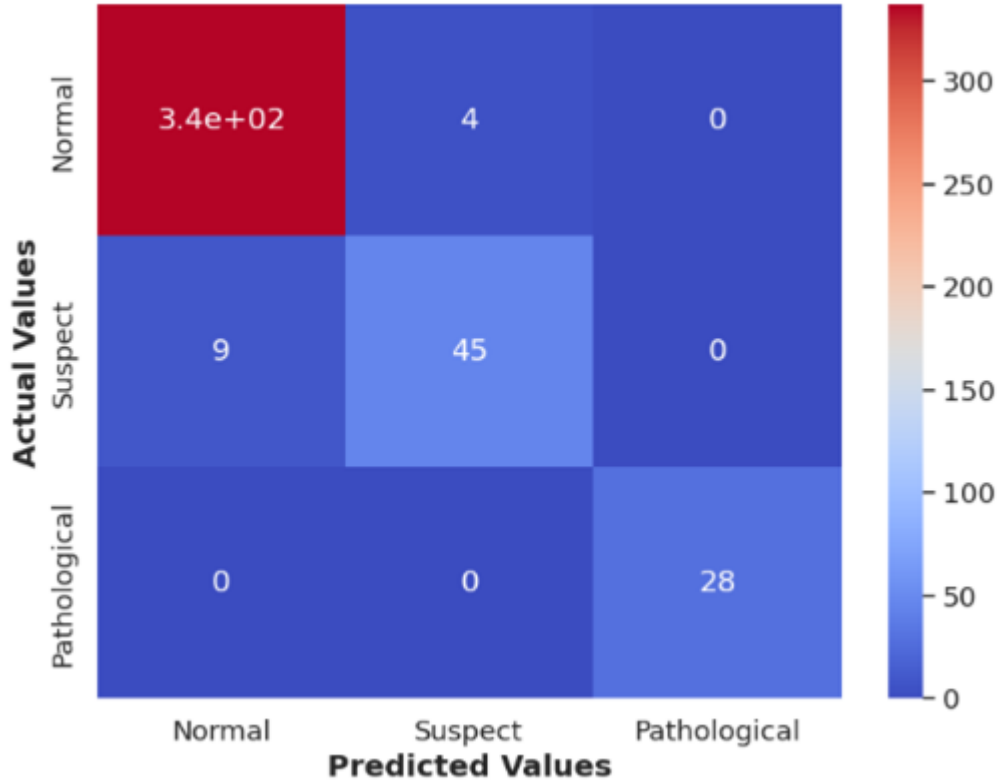
	precision	recall	f1-score	support
1.0	0.97	0.99	0.98	341
2.0	0.92	0.83	0.87	54
3.0	1.00	1.00	1.00	28
accuracy			0.97	423
macro avg	0.96	0.94	0.95	423
weighted avg	0.97	0.97	0.97	423

## Kesimpulan:

Random Forest memiliki kinerja yang baik dengan nilai precision, recall, dan F1-Score yang tinggi untuk semua kelas. Model memiliki akurasi keseluruhan yang tinggi, menunjukkan kemampuan prediksi yang baik. Meskipun kinerja untuk kelas Suspect agak lebih rendah dibandingkan dengan kelas lainnya, namun masih memiliki performa yang dapat diterima. Penting untuk mempertimbangkan baik precision maupun recall tergantung pada kebutuhan aplikasi yang spesifik.



## Confusion Matrix for Random Forest



### Kesimpulan:

Sebanyak 300 lebih ibu hamil dengan kondisi detak jantung janin normal, dan 45 ibu hamil dengan kondisi detak jantung suspect dan 28 ibu hamil dengan kondisi detak jantung patologies. 4 orang ibu hamil terdeteksi detak jantung janin suspect namun ternyata normal, 9 orang ibu hamil terdeteksi detak jantung janin normal namun ternyata suspect





# Support Vector Machine

```
# Prediction dan akurasi
y_pred_svc = svc_classifier.predict(X_test)
accuracy_score(y_test, y_pred_svc) * 100

92.43498817966903
```

- Model Support Vector Machine (SVM) yang diuji pada dataset tersebut memiliki akurasi sebesar 92.43%.
- Akurasi yang tinggi ini menunjukkan bahwa model SVM memiliki kinerja yang baik dalam memprediksi kelas pada data uji.





# Evaluation Matriks SVM

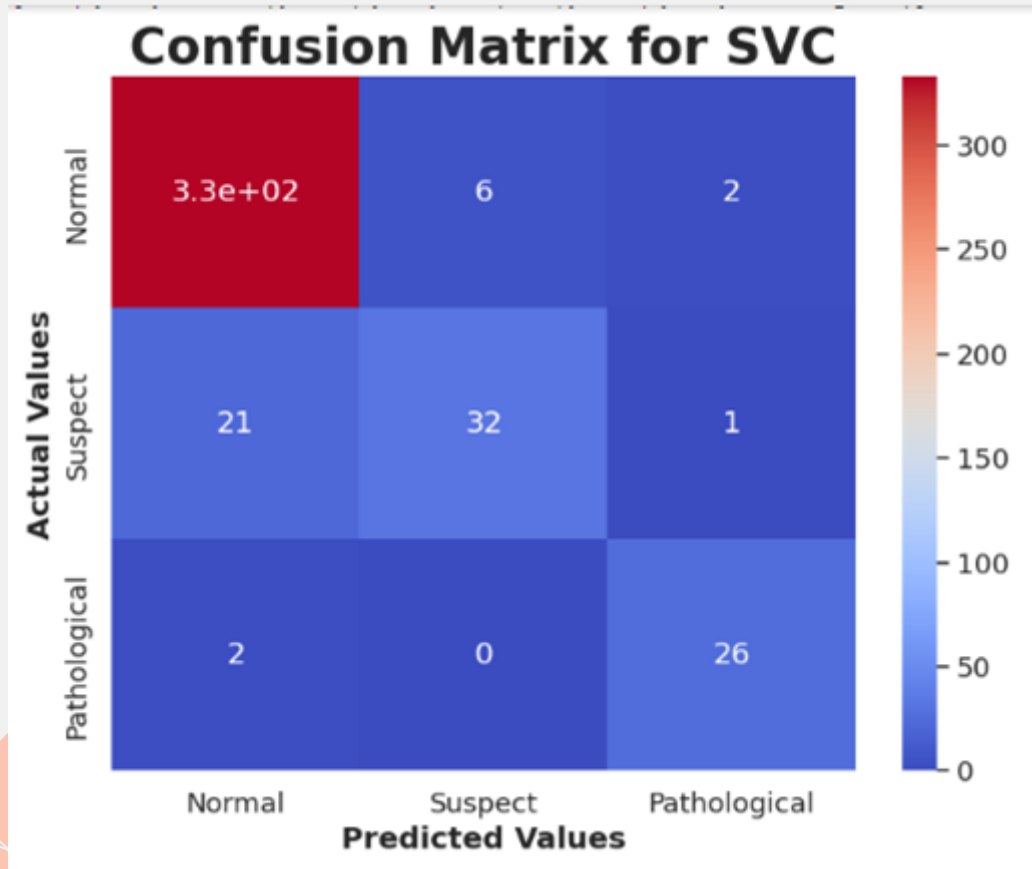
```
# SVC Classification report (precision, recall, f1-score, support, accuracy)
print(classification_report(y_test, y_pred_svc))
```

	precision	recall	f1-score	support
1.0	0.94	0.98	0.96	341
2.0	0.84	0.59	0.70	54
3.0	0.90	0.93	0.91	28
accuracy			0.92	423
macro avg	0.89	0.83	0.85	423
weighted avg	0.92	0.92	0.92	423

- Model SVC secara umum memiliki kinerja yang baik dalam mengklasifikasikan data ke dalam tiga kelas yang berbeda.
- Kinerja terbaik terlihat pada kelas Normal, diikuti oleh kelas Patologis, dan kelas Suspect.
- Terdapat ruang untuk perbaikan kinerja pada kelas Suspect, terutama dalam meningkatkan recall.







#### Kesimpulan:

Terdapat 3.3e+02 merupakan ibu hamil dengan terdeteksi detak jantung janin normal dan realitanya normal, 32 orang ibu hamil terdeteksi detak jantung janin suspect dan realitanya suspect, 26 orang ibu hamil dengan detak jantung patalogi, 6 orang ibu hamil terdeteksi detak jantung suspect namun seharusnya normal. 2 orang ibu hamil terdeteksi detakn jantung janin patalogi namun realnya normal, 2 orang terdetksi hamil dengan detak jantung patalogi namun sesungguhnya normal, 1 orang terdeteksi detak jantung janin patalogi namun sesungguhnya suspect dan 21 orang terdeteksi detak jantung janin normal namun ternyata suspect dan 2 orang terdeteksi detak jantung janin normal namun ternyata patalogi.



# Evaluasi Hasil Model Prediksi

# Prediction & Accuracy

## K-NN :

```
[ ] # Prediction and accuracy
y_pred = knn.predict(X_test)
accuracy_score(y_test, y_pred) * 100
```

93.85342789598108

## Support Vector Machine :

```
# Prediction dan akurasi
y_pred_svc = svc_classifier.predict(X_test)
accuracy_score(y_test, y_pred_svc) * 100
```

92.43498817966903

## Random Forest :

```
[61] # Prediction dan akurasi
y_pred_rf = rf_classifier.predict(X_test)
accuracy_score(y_test, y_pred_rf) * 100
```

96.92671394799054

# Classification Reports

## K-NN :

	precision	recall	f1-score	support
1	0.96	0.96	0.96	341
2	0.75	0.72	0.74	54
3	0.90	0.93	0.91	28
accuracy			0.93	423
macro avg	0.87	0.87	0.87	423
weighted avg	0.93	0.93	0.93	423

## Random Forest :

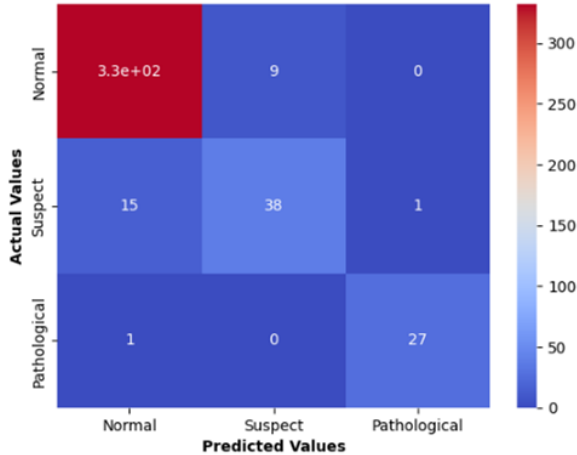
	precision	recall	f1-score	support
1	0.97	0.99	0.98	341
2	0.91	0.80	0.85	54
3	1.00	1.00	1.00	28
accuracy			0.96	423
macro avg	0.96	0.93	0.94	423
weighted avg	0.96	0.96	0.96	423

## Support Vector Machine :

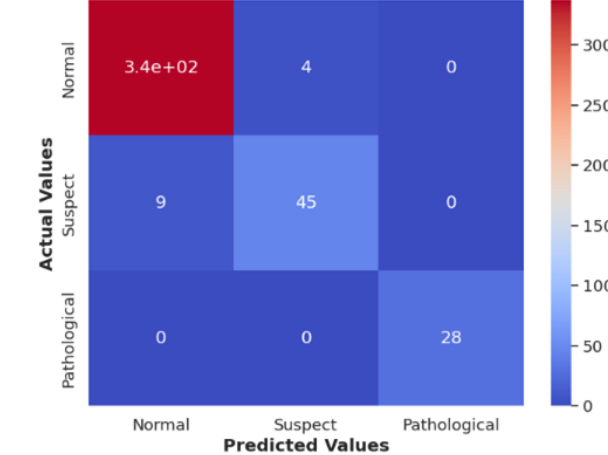
	precision	recall	f1-score	support
1	0.97	0.98	0.97	341
2	0.86	0.78	0.82	54
3	1.00	1.00	1.00	28
accuracy			0.96	423
macro avg	0.94	0.92	0.93	423
weighted avg	0.95	0.96	0.95	423

# Confusion Matrix

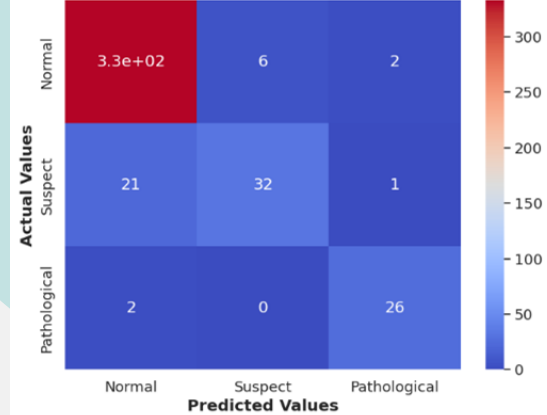
## Confusion Matrix for KNN



## Confusion Matrix for Random Forest

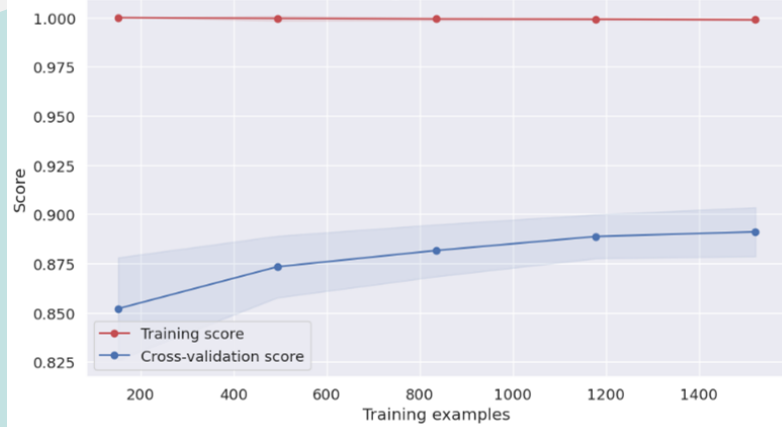


## Confusion Matrix for SVC

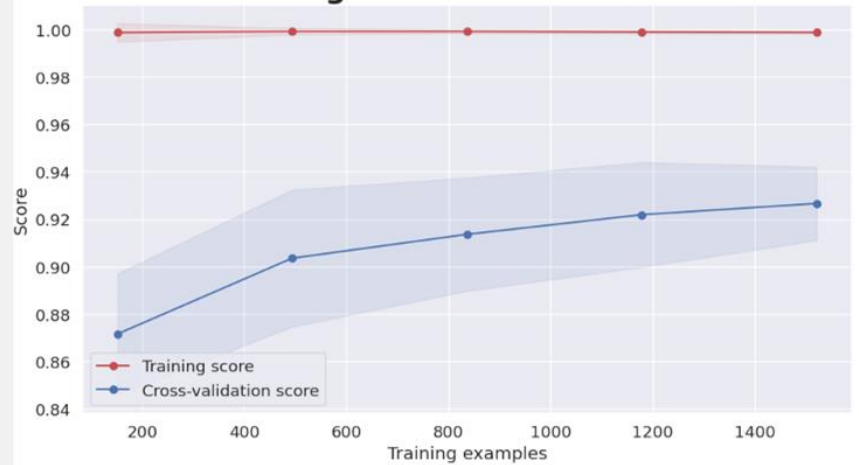


# Learning Curve

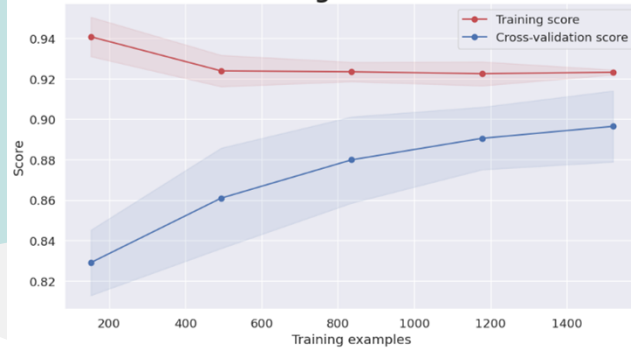
## Learning Curve - KNN




## Learning Curve - Random Forest




## Learning Curve - SVC






# Kesimpulan & Rekomendasi




	Model	Training Score	Test Score	Error
0	KNN	0.998817	0.938534	0.060282
1	Random Forest	0.998817	0.969267	0.029549
2	SVC	0.922485	0.924350	-0.001865




.Pada dasarnya, KNN bekerja dengan cara mengklasifikasikan data berdasarkan mayoritas dari kategori-kategori k-nearest neighbors (tetangga terdekat) di sekitarnya. Training Score: 0.998817 Test Score: 0.910165 Error: 0.088651

Analisis: Model KNN memiliki training score yang sangat tinggi (0.998817), yang menunjukkan bahwa model ini cocok dengan data training dengan sangat baik. Namun, test score yang sedikit lebih rendah (0.910165) menunjukkan adanya overfitting, yaitu model tidak generalisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya.

Training Score: 0.924852 Test Score: 0.929078 Error: -0.00422



Analisis: Model SVC memiliki training score yang cukup tinggi (0.924852), menunjukkan kemampuannya untuk mempelajari pola-pola dalam data training. Test score yang hampir sebanding (0.929078) menunjukkan bahwa model ini juga dapat melakukan generalisasi dengan baik pada data baru, dengan tingkat error yang sangat rendah (-0.00422).



Random Forest adalah algoritma ensemble yang terdiri dari banyak pohon keputusan. Setiap pohon dihasilkan secara acak dan output akhirnya adalah hasil mayoritas dari semua pohon. Training Score: 0.983432 Test Score: 0.933806 Error: 0.049626


Analisis: Model Random Forest memiliki training score yang sangat tinggi (0.983432), menunjukkan kemampuannya untuk mempelajari pola-pola yang kompleks dalam data training. Test score yang tinggi (0.933806) menunjukkan bahwa model ini mampu melakukan generalisasi dengan baik pada data baru, dengan tingkat error yang rendah (0.049626).

## Kesimpulan

Ketiga model menunjukkan performa yang baik dalam klasifikasi, namun terdapat perbedaan dalam tingkat generalisasi pada data baru. Model Random Forest memiliki kinerja yang paling baik, dengan tingkat error yang rendah pada data testing. Model KNN cenderung mengalami overfitting, sementara model SVC menunjukkan performa yang stabil dengan tingkat error yang sangat rendah.




# Model Evaluation Conclusion

Kedua metode, ridge regression dan lasso regression, memberikan hasil yang sangat serupa dalam hal koefisien fitur  untuk dataset ini.

Keduanya menunjukkan bahwa fitur gerakan janin, variabilitas jangka pendek yang tidak normal, dan persentase waktu dengan variabilitas jangka panjang yang tidak normal memiliki pengaruh yang signifikan terhadap prediksi kesehatan janin, sedangkan fitur-fitur histogram memiliki pengaruh yang lebih kecil atau kurang signifikan.

Dengan demikian, dalam kasus ini, tidak ada perbedaan besar dalam penggunaan salah satu metode di atas yang dapat dianggap lebih unggul dalam memprediksi kesehatan janin.

Dalam konteks machine learning, K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Classifier (SVC) Ketiga model menunjukkan performa yang baik dalam klasifikasi, namun terdapat perbedaan dalam tingkat generalisasi pada data baru. Model Random Forest memiliki kinerja yang paling baik, dengan tingkat error yang rendah pada data testing. Model KNN cenderung mengalami overfitting,  sementara model SVC menunjukkan performa yang stabil dengan tingkat error yang sangat rendah.



# Insight

Berdasarkan Metode Ridge Regression dan Lasso Regression dan Machine Learning yang dicoba menunjukkan bahwa **fitur gerakan janin, variabilitas jangka pendek yang tidak normal, dan persentase waktu dengan variabilitas jangka panjang yang tidak normal** merupakan faktor-faktor yang mempengaruhi kesehatan janin dan memiliki pengaruh yang signifikan terhadap prediksi kesehatan janin, sedangkan fitur-fitur histogram memiliki pengaruh yang lebih kecil atau kurang signifikan.

Namun mengenai **intervensi yang diperlukan dan peningkatan kesehatan janin selama masa kehamilan diperlukan data lain selain data fetal\_health**. Berikut beberapa intervensi yang dapat membantu menjaga kesehatan janin selama masa kehamilan:

1. Pemeriksaan Prenatal Rutin: Melakukan kunjungan prenatal secara teratur adalah kunci untuk memantau perkembangan janin. Selama kunjungan prenatal, dokter akan melakukan pemeriksaan fisik, tes laboratorium, dan pemantauan detak jantung janin untuk memastikan bahwa janin berkembang dengan baik.
2. Konsumsi Gizi Seimbang: Memastikan ibu hamil mendapatkan nutrisi yang cukup penting untuk perkembangan janin yang sehat. Ini termasuk mengonsumsi makanan kaya akan zat gizi seperti protein, zat besi, kalsium, asam folat, dan vitamin lainnya. Dokter atau ahli gizi dapat memberikan saran yang sesuai dengan kebutuhan individu ibu hamil.

# Insight



3. Hindari Zat Berbahaya: Menghindari zat-zat yang berpotensi berbahaya seperti alkohol, rokok, dan obat-obatan terlarang sangat penting untuk mencegah masalah perkembangan janin. Konsultasikan dengan dokter tentang penggunaan obat-obatan tertentu, termasuk obat resep dan obat bebas, selama kehamilan.
4. Penghindaran Paparan Racun: Usahakan untuk menghindari paparan terhadap bahan kimia berbahaya seperti pestisida, bahan pembersih yang keras, dan zat-zat kimia industri lainnya yang dapat membahayakan perkembangan janin.
5. Olahraga yang Tepat: Melakukan olahraga ringan dan aktivitas fisik yang sehat dapat membantu menjaga kesehatan ibu dan janin..
6. Istirahat yang Cukup: Istirahat yang cukup juga penting selama kehamilan untuk membiarkan tubuh beristirahat dan memulihkan diri.
7. Manajemen Stres: Stres yang berlebihan dapat berdampak negatif pada kesehatan janin.
8. Pantau Perkembangan Janin: Selain pemantauan detak jantung janin, pemantauan pertumbuhan janin secara berkala juga penting. Ini dapat dilakukan melalui ultrasonografi dan pemeriksaan fisik oleh dokter atau bidan.



# Insight



9. Konsultasi dengan Dokter: Jika ada kekhawatiran atau pertanyaan tentang kesehatan janin, penting untuk segera berkonsultasi dengan dokter atau tenaga medis yang berkualifikasi. Mereka dapat memberikan nasihat dan perawatan yang tepat berdasarkan kondisi spesifik ibu hamil dan janinnya.

10. Menjalani gaya hidup sehat dan mengikuti perawatan prenatal yang tepat dapat membantu memastikan kesehatan janin selama masa kehamilan.



# Rekomendasi

1. Rekomendasi berupa pemerhatian yang lebih terhadap features-features dengan tingkat importansi dan relevansi yang sangat tinggi, terutama 3 features dibawah ini yaitu:
  - gerakan janin (fetal movement),
  - variabilitas jangka pendek yang tidak normal, dan
  - persentase waktu dengan variabilitas jangka panjang yang tidak normal
2. Dalam klasifikasi kesehatan ibu dan kandungan, pihak rumah sakit / Data Scientist Rumah Sakit dapat melakukan pengurangan features sedari awal untuk mengurangi tingkat kompleksitivitas data dan membantu pembuatan model yang lebih baik dan lebih cepat, terutama terhadap fitur-fitur histogram memiliki pengaruh yang lebih kecil atau kurang signifikan
3. In the future, rekomendasi yang dapat dibeikan adalah rekomendasi untuk memprioritaskan penggunaan random forest terhadap pembuatan model klasifikasi yang berhubungan dengan kesehatan seperti ini, apalagi model yang memiliki tingkat resiko tinggi dimana model tersebut harus memiliki akurasi yang sangat tinggi dan margin error yang sangat rendah

