**MENGANALISA RESIKO MANTAN NAPI TERTANGKAP KEMBALI DENGAN MODEL COX PROPORTIONAL HAZARD**

Muhammad Fakhry Firdausi,1806208213 Muhammad Satria Ibrahim, 1806208062 Muhammad Mirza Iskandar, 1806193413

Email:

[muhammad.fakhry@sci.ui.ac.id,](mailto:muhammad.fakhry@sci.ui.ac.id) [muhammad.satria83@sci.ui.ac.id,](mailto:muhammad.satria83@sci.ui.ac.id) [muhammad.mirza81@sci.ui.ac.id,](mailto:muhammad.mirza81@sci.ui.ac.id)

## Pendahuluan

Pada tahun 1970-an, 432 narapidana yang dibebaskan dari penjara negara bagian Maryland Amerika Serikat. Sebagian dari mereka secara acak diberikan bantuan keuangan dan setengah yang lain tidak diberikan bantuan keuangan sebagai bagian dari sebuah eksperimen. Kemudian mereka ditindaklanjuti dan didata kembali satu tahun setelah mereka dibebaskan.

Eksperimen tersebut bertujuan untuk melihat berapa banyak narapidana yang kembali dipenjarakan setelah dibebaskan dan faktor-faktor pendukungnya. Hal ini tentu penting untuk dipelajari, karena hal tersebut dapat berpengaruh besar dalam bertambahnya biaya substansial pada masyarakat baik secara langsung maupun tidak langsung, seperti biaya penahanan, biaya keadilan dan aparat polisi, kurangnya produksi, dan lain sebagainya.

Tujuan dari analisis ini adalah untuk mengolah, mengevaluasi, dan memberi saran serta rekomendasi terkait hasil data dari eksperimen tersebut. Untuk itu ada beberapa pertanyaan yang akan dijawab pada analisis ini :

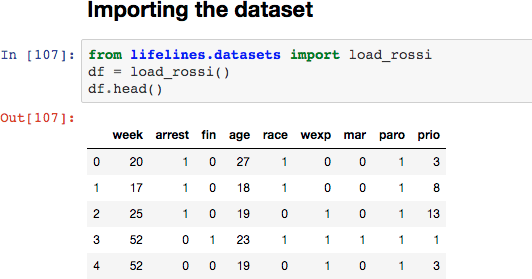
* Faktor-faktor apa saja yang mempengaruhi resiko mantan narapidana akan tertangkap lagi?
* Seberapa besar pengaruh faktor-faktor terhadap fungsi hazard dari mantan narapidana?

## Data

Data yang kami gunakan dalam analisis ini adalah data narapidana yang dibebaskan dari penjara negara bagian Maryland di Amerika Serikat. Terdapat 432 narapidana yang dibebaskan pada tahun 1970-an. Setengah dari narapidana tersebut diberikan bantuan keuangan, sedangkan setengah yang lain tidak. Kemudian mereka kembali didata setahun setelah mereka dibebaskan dari penjara. Selain itu variabel-variabel lain yang ikut didata antara lain adalah : umur, ras, pengalaman kerja, status pernikahan, banyaknya hukuman sebelum penahanan kembali, bebas bersyarat atau tidak, dalam penahanan atau tidak, dan waktu dalam minggu dari pembebasan hingga penahanan kembali.

Sumber data *:* [*https://lifelines.readthedocs.io/en/latest/lifelines.datasets.html?highlight=datasets*](https://lifelines.readthedocs.io/en/latest/lifelines.datasets.html?highlight=datasets)

|  |  |
| --- | --- |
| Variabel | Keterangan |
| week | waktu dalam minggu dari pembebasan hingga penahanan kembali |
| arrest | penahanan kembali (1 = ya, 0 = tidak) |
| fin | bantuan keuangan (1 = ya, 0 = tidak) |
| age | usia narapidana |
| race | ras narapidana (1 = kulit hitam, 0 = bukan kulit hitam) |
| wexp | pengalaman kerja sebelum penahanan (1 = ya, 0 = tidak) |
| mar | status pernikahan (1 = ya, 0 = tidak) |
| paro | bebas bersyarat (1 = ya, 0 = tidak) |
| prio | banyaknya hukuman sebelum penahanan kembali |



#### Bagaimana cara membaca data tersebut?

Narapidana kedua (baris kedua) tertangkap kembali (arrest = 1) 20 minggu (week = 20) setelah dibebaskan, tidak mendapatkan bantuan finansial (fin = 0), berusia 18 tahun (age = 18), kulit hitam (race = 1), tidak mempunyai pengalam kerja sebelum penahanan (wexp = 0), belum menikah (mar = 0), bebas bersyarat (paro = 1), mempunyai 8 hukuman sebelum penahanan kembali (prio = 8).

Narapidana ke 4 (baris keempat) tidak tertangkap kembali (arrest = 0) 52 minggu (week = 52) setelah dibebaskan, mendapatkan bantuan finansial (fin = 1), berusia 23 tahun (age = 23), kulit hitam (race = 1), mempunyai pengalam kerja sebelum penahanan (wexp = 1), sudah menikah (mar = 1), bebas bersyarat (paro = 1), mempunyai 1 hukuman sebelum penahanan kembali (prio = 1).

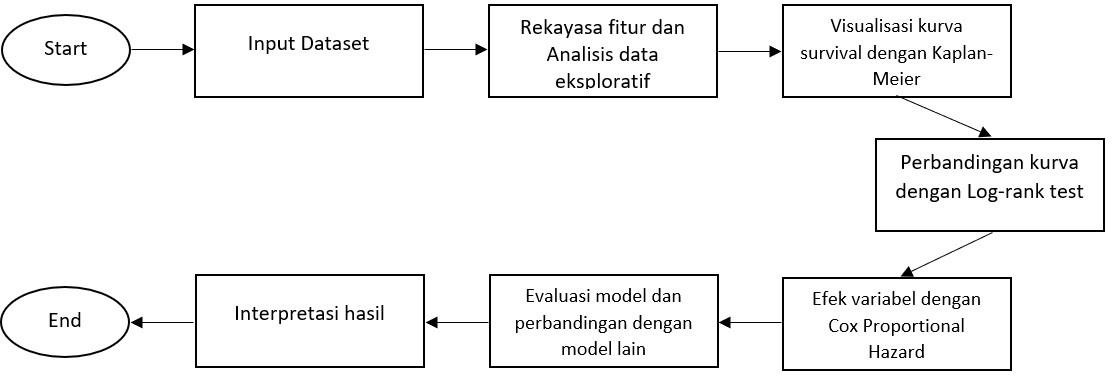
#### Apa yang dapat dianalisis pada data ini?

Berikut beberapa hal yang dapat dielaborasi dari data ini :

* Faktor-faktor apa saja yang mempengaruhi resiko mantan narapidana akan tertangkap lagi?
* Seberapa besar pengaruh faktor-faktor terhadap fungsi hazard dari mantan narapidana?

# Pengolahan data dan interpretasi

Berikut langkah-langkah pengolahan data yang kami lakukan pada analisis ini :



Evaluasi Model

### Import Package ke Notebook



1. **Import Dataset**

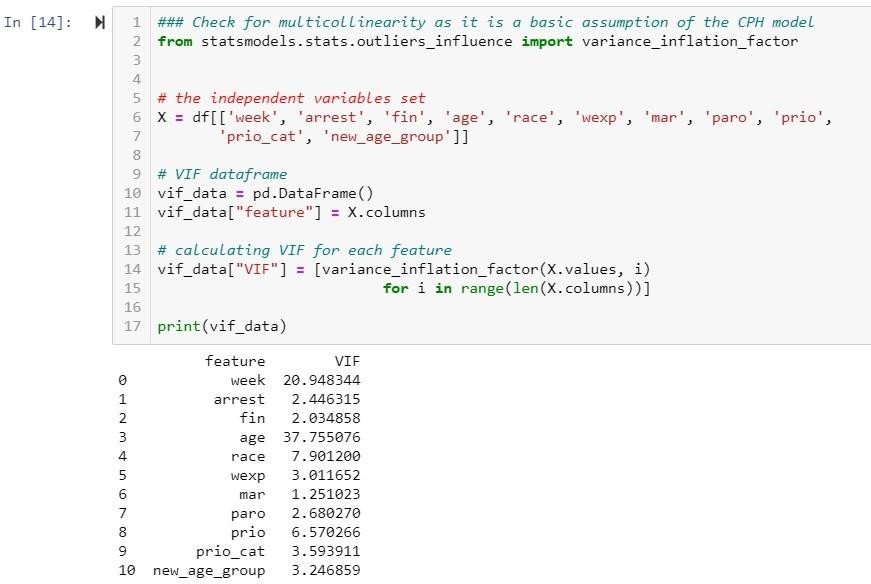


1. **Feature Engineering**

Akan dilakukan pengecekan multikolinearitas, karena model coxPH akan terpengaruhi jika ada fitur fitur yang multikolinear. Akan dilakukan pengecekan dengan *Variance Inflation Factor*.

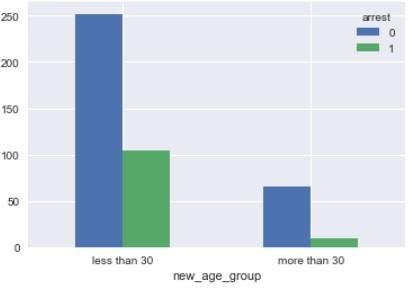


Karena terdapat multikolinearitas yang tinggi antara fitur **week** dan fitur **age**, maka akan dilakukan *binning* untuk fitur age dengan mengelompokkan data, akan bernilai 0 jika berusia kurang dari 30 tahun, dan akan bernilai 1 jika berusia lebih dari 30 tahun.

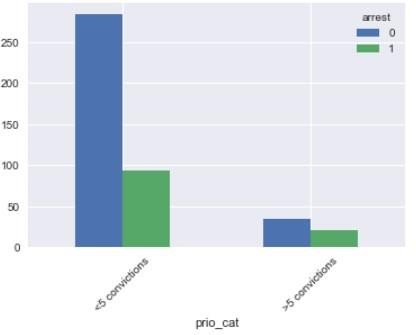
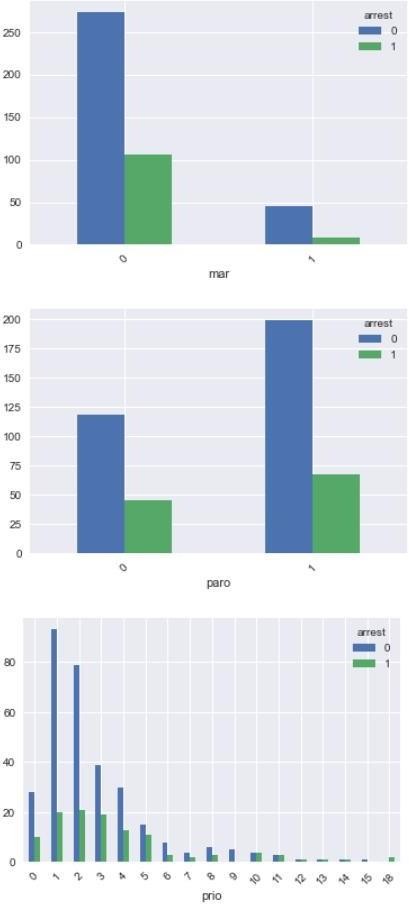
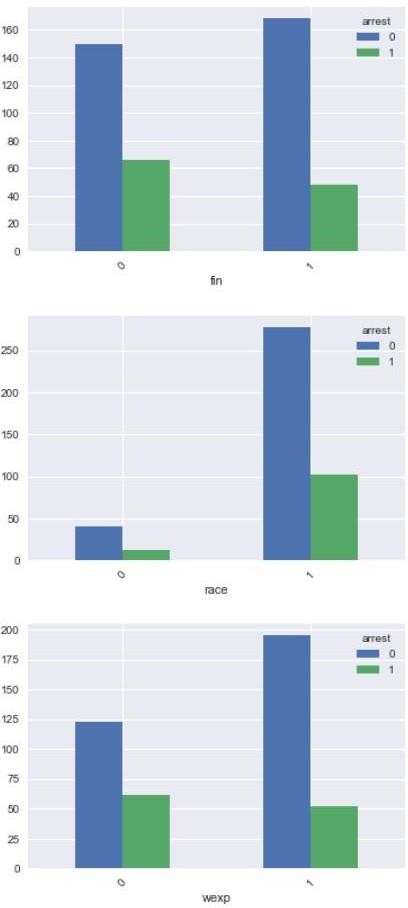


Setelah dilakukan binning, berdasarkan pengecekan menggunakan ***VIF***, didapati fitur *new\_age\_group* sudah aman dari multikolinearitas, maka akan kita drop fitur age. Dan untuk berjaga jaga, kami juga melakukan binning pada fitur prio.

### Exploratory Data Analysis



Didapati wawasan bahwa persentase untuk mantan napi yang tertangkap lagi dari kelompok usia kurang dari 30 tahun yakni sebesar 29%. Sedangkan persentase untuk mantan napi yang tertangkap lagi dari kelompok usia kurang dari 30 tahun yakni sebesar 12%.



Dari visualisasi *categorical features,* kita dapati beberapa wawasan sebagai berikut:

* Mantan napi yang tanpa bantuan keuangan, lebih banyak dilakukan pengangkapan kembali.
* Mantan napi yang berkulit hitam lebih banyak dilakukan penangkapan kembali dibanding ras non-hitam.
* Mantan napi yang tidak mempunyai pengalaman kerja sebelumnya, lebih banyak dilakukan penangkapan kembali.
* Mantan napi yang belum menikah lebih banyak dilakukan penangkapan kembali, bisa jadi dikarenakan untuk yang sudah berkeluarga akan lebih mempertimbangkan untuk melakukan tindakan kriminalitas kembali karena mempunyai tanggungan keluarga.
* Mantan napi yang bebas bersyarat justru lebih banyak dilakukan penangkapan kembali.
* Mantan napi yang punya jumlah catatan criminal kurang dari 5 kali lebih banyak dilakukan penangkapan kembali.

### Kaplan Meier Curves

Dengan tanpa mempertimbangkan masing-masing fitur, berdasarkan keseluruhan data, didapati fungsi survival dari *Kurva Kaplan Meier* untuk data yang diamati sebagai berikut:

Dapat kita lihat bahwa berdasarkan fungsi survival dari Kurva Kaplan-Meier yang didapati, bahwa dalam 10 minggu pertama, telah terjadi pengangkapan sebesar 4% dari seluruh populasi mantan napi, dan setelah 40 minggu, didapati bahwa telah terjadi penurunan fungsi survival sebesar 20%. Ini berarti bahwa setelah 40 minggu, telah dilakukan penangkapan kembali mantan napi sebanyak 20% dari seluruh populasi.

Kita juga dapati wawasan bahwa dalam interval waktu 0 hingga 52 minggu, jumlah mantan napi yang tertangkap lagi sejumlah 114, sedangkan sisanya merupakan termasuk data tersensor kanan.

*Akan dilihat estimasi fungsi survival untuk masing-masing fitur dengan menggunakan Kurva Kaplan Meier.*

#### Fitur race

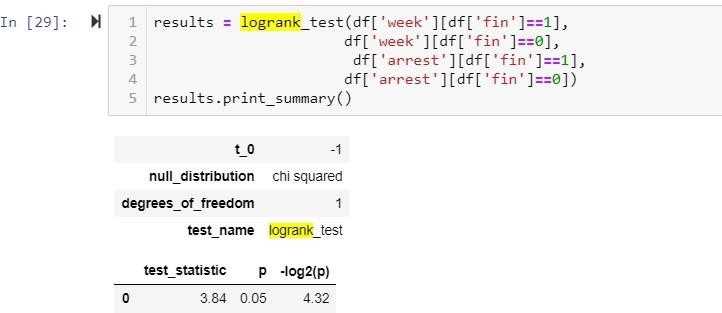


Didapati bahwa untuk populasi mantan napi berkulit hitam, mengalami penurunan fungsi survival yang lebih signifikan dibanding mantan napi berkulit non-hitam. Ini berarti bahwa mantan napi berkulit hitam lebih beresiko dalam melakukan tindakan kriminal kembali.

#### Fitur Financial Aid



Didapati bahwa dari estimasi, fungsi survival untuk mantan napi yang memiliki bantuan keuangan terjadi penurunan yang lebih signifikan dari yang tidak punya bantuan keuangan. Namun dapat dilihat dari kurva Kaplan Meier diatas, kedua kurva tersebut berhimpitan. Maka akan dilakukan Uji logrank untuk uji hipotesis melihat apakah ada pengaruh yang berbeda diantara kedua fitur tersebut.



karena didapati p-value yang sama dengan taraf signifikansi, maka dapat disimpulkan mantan napi yang **Memiliki Bantuan Keuangan** dan **Tidak Memiliki Bantuan Keuangan** , akan **memberikan efek resiko tertangkap yang berbeda.**

#### Fitur Is Married

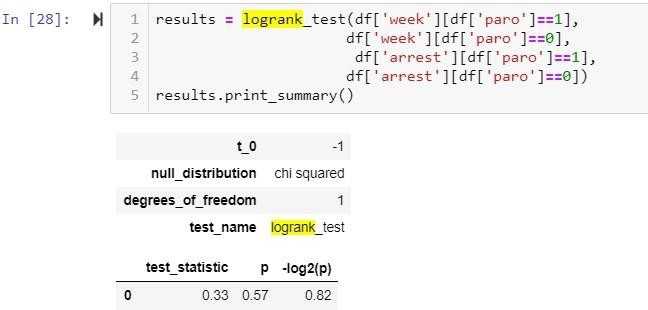


Didapati bahwa untuk mantan napi yang belum menikah, mengalami penurunan fungsi survival yang lebih jauh dibanding mantan napi yang sudah menikah. Bisa jadi dipengaruhi oleh banyaknya tanggunan keluarga.

#### Fitur Released On Parole

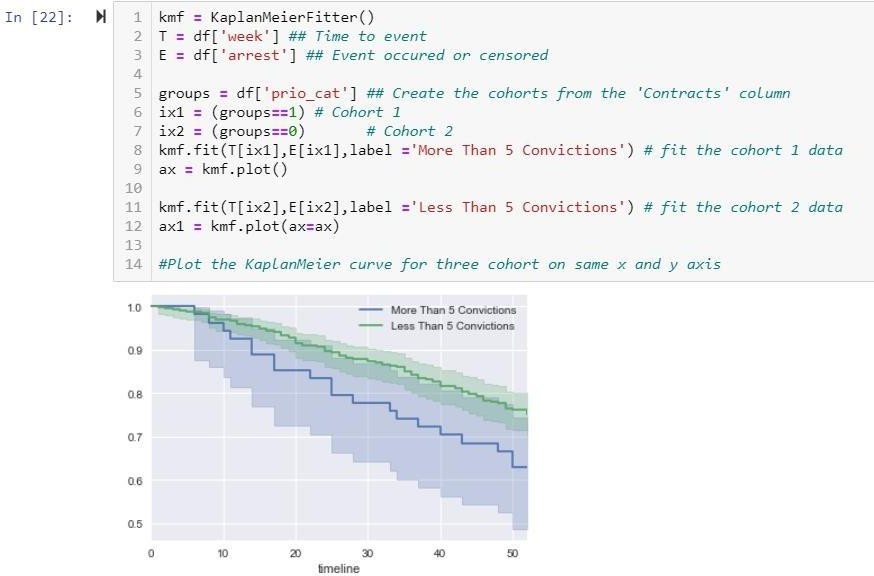


Didapati perbandingan fungsi survival dari mantan napi yang *Bebas Bersyarat* dan yang *Bebas Tidak Bersyarat* sangatlah serupa. Maka akan dilakukan Uji logrank untuk melihat apakah ada pengaruh yang berbeda diantara kedua ciri tersebut.



Setelah melakukan Uji logrank, didapati **p-value** yang lebih besar dari taraf signifikansi, ini berarti jika mantan napi *Bebas Bersyarat* maupun *Bebas Tidak Bersayart* akan memberikan efek yang sama pada fungsi survival.

#### Fitur Prio (Number of Convictions)



Didapati bahwa fungsi survival untuk mantan napi yang memiliki lebih dari 5 catatan kriminal mengalami penurunan yang lebih signifikan dibanding mantan napi yang memiliki kurang dari 5 catatan kriminal.

#### Fitur new\_age\_group



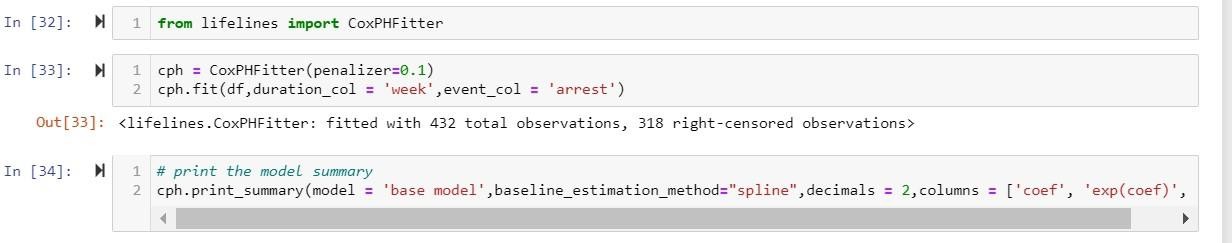
Didapati bahwa fungsi survival untuk mantan napi yang berusia kurang dari 30 tahun mengalami penurunan fungsi survival yang lebih signifikan dibanding mantan napi yang berusia lebih dari 30 tahun.

### Encoding the data

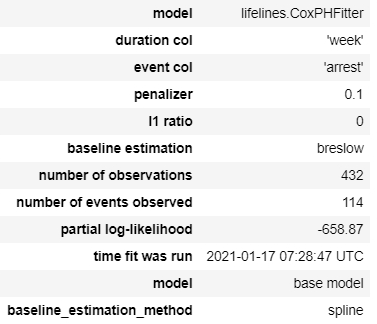
Sebelum dilakukan permodelan fungsi *Cox proportional hazard*, akan dilakukan Data Encoding untuk membuat dummy variables. Disini untuk fitur yang akan dibuat dummy variable nya yakni fitur ***prio***.

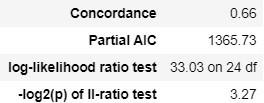


### Cox Proportional Hazard Modelling

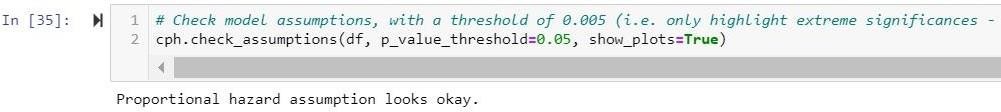


Digunakan package lifelines untuk mengimport model CoxPHFitter. Dengan menggunakan syntax

*cph.print\_summary()* akan kita dapati hasil sebagai berikut:

Dari hasil diatas, didapati Concordance Index yakni sebesar 0.66 yang berarti mutu dari model yang kita dapati yakni 66% dan angka tersebut sudah termasuk nilai yang dapat diterima.

Akan dicek asumsi dari model yang kita gunakan, untuk melihat apakah dari fitur fitur yang kita gunakan dapat ditarik kesimpulan.



Setelah melakukan pengecekan asumsi, dengan menetapkan *p-value threshold* sebesar 0.05, didapati bahwa model yang kita gunakan sudah cukup baik, berarti setiap fitur dari model kita dapat ditarik kesimpulannya. Maka, berdasarkan tabel summary yang kita dapati, dapat kita interpretasikan wawasan wawasan mengenai resiko dari masing masing fitur terhadap fungsi hazard sebagai berikut;

* Untuk fitur *financial aid* atau bantuan keuangan, didapat bahwa untuk mantan narapidana yang tidak mendapatkan bantuan keuangan lebih beresiko sebesar 128% untuk tertangkap kembali dibandingkan dengan yang mendapatkan bantuan keuangan.
* Untuk fitur *race* atau ras, didapat bahwa untuk mantan narapidana yang tidak berkulit putih lebih beresiko sebesar 122% untuk tertangkap kembali dibandingkan dengan yang berkulit putih.
* Untuk fitur *work experience* atau pengalaman kerja, didapat bahwa untuk mantan narapidana yang mempunyai pengalaman kerja sebelum penahanan lebih beresiko sebesar 128% untuk tertangkap kembali dibandingkan dengan yang tidak mempunyai pengalaman kerja.
* Untuk fitur *marital status* atau status pernikahan, didapat bahwa untuk mantan narapidana yang belum menikah lebih beresiko sebesar 142% untuk tertangkap kembali dibandingkan dengan yang sudah menikah.
* Untuk fitur *parole* atau bebas bersyarat, dari sebelumnya sudah didapat bahwa untuk mantan narapidana yang bebas bersyarat maupun tidak akan memberikan efek yang sama pada fungsi survival. Sehingga fitur ini tidak mempunyai pengaruh yang signifikan dalam resiko penahanan kembali.
* Untuk fitur *prior convictions category* atau kategori banyaknya hukuman sebelum penahanan, didapat bahwa untuk mantan narapidana yang mempunyai lebih dari 5 hukuman lebih beresiko sebesar 115% untuk tertangkap kembali dibandingkan dengan yang mempunyai kurang dari 5 hukuman.
* Untuk fitur *new age group* atau grup usia, didapat bahwa untuk mantan narapidana yang berusia kurang dari 30 tahun lebih beresiko sebesar 151% untuk tertangkap kembali dibandingkan dengan yang berusia lebih dari 30 tahun.
* Untuk fitur *n-prior convictions* atau n-hukuman sebelum penahanan, didapat bahwa untuk mantan narapidana yang mempunyai hukuman sebanyak : 0, 3, 4, 5, 7, 10, 11, 12, 13, 14, 18 kali lebih beresiko untuk tertangkap kembali dibandingkan yang lainnya.

# Kesimpulan

Maka berdasarkan analisa yang kami dapati dari data yang kami amati, didapati kesimpulan untuk menjawab pertanyaan berikut:

* + Faktor-faktor apa saja yang mempengaruhi resiko mantan narapidana akan tertangkap lagi?

Berdasarkan Analisa yang kami lakukan, didapati bahwa fitur-fitur yang berpengaruh terhadap fungsi hazard dari data yang kami amati adalah semua fitur terkecuali untuk fitur **paro**.

* + Seberapa besar pengaruh faktor-faktor terhadap fungsi hazard dari mantan narapidana?
* Untuk fitur *financial aid* atau bantuan keuangan, didapat bahwa untuk mantan narapidana yang tidak mendapatkan bantuan keuangan lebih beresiko sebesar 128% untuk tertangkap kembali dibandingkan dengan yang mendapatkan bantuan keuangan.
* Untuk fitur *race* atau ras, didapat bahwa untuk mantan narapidana yang berkulit hitam lebih beresiko sebesar 122% untuk tertangkap kembali dibandingkan dengan yang bukan berkulit hitam.
* Untuk fitur *work experience* atau pengalaman kerja, didapat bahwa untuk mantan

narapidana yang tidak mempunyai pengalaman kerja sebelum penahanan lebih beresiko sebesar 128% untuk tertangkap kembali dibandingkan dengan yang mempunyai pengalaman kerja.

* Untuk fitur *marital status* atau status pernikahan, didapat bahwa untuk mantan narapidana yang belum menikah lebih beresiko sebesar 142% untuk tertangkap kembali dibandingkan dengan yang sudah menikah.
* Untuk fitur *parole* atau bebas bersyarat, dari sebelumnya sudah didapat bahwa untuk mantan narapidana yang bebas bersyarat maupun tidak akan memberikan efek yang sama pada fungsi survival. Sehingga fitur ini tidak mempunyai pengaruh yang signifikan dalam resiko penahanan kembali.
* Untuk fitur *prior convictions category* atau kategori banyaknya hukuman sebelum penahanan, didapat bahwa untuk mantan narapidana yang mempunyai lebih dari 5 hukuman lebih beresiko sebesar 115% untuk tertangkap kembali dibandingkan dengan yang mempunyai kurang dari 5 hukuman.
* Untuk fitur *new age group* atau grup usia, didapat bahwa untuk mantan narapidana yang berusia kurang dari 30 tahun lebih beresiko sebesar 151% untuk tertangkap kembali dibandingkan dengan yang berusia lebih dari 30 tahun.
* Untuk fitur *n-prior convictions* atau n-hukuman sebelum penahanan, didapat bahwa untuk mantan narapidana yang mempunyai hukuman sebanyak : 0, 3, 4, 5, 7, 10, 11, 12, 13, 14, 18 kali lebih beresiko untuk tertangkap kembali dibandingkan yang lainnya.

Untuk narapidana dengan 3 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**122%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 4 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**122%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 5 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**159%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 10 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**164%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 11 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**186%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 12 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**166%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 13 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**171%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 14 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**198%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Untuk narapidana dengan 18 hukuman sebelum penahanan, didapati bahwa lebih beresiko sebesar

**651%** untuk tertangkap kembali dibandingkan dengan yang lainnya.

Berdasarkan hasil analisa yang kami dapati, dapat kami rekomendasikan ke pihak kepolisian bahwa mantan narapidana yang memilki salah satu atau lebih dari ciri ciri sebagai berikut: *Tidak mempunyai bantuan keuangan, berkulit hitam, tidak punya pengalaman kerja, berusia kurang dari 30 tahun,* dan *memiliki jumlah catatan kriminal sebanyak 3,4,5,10,11,12,13,14 dan 18 kali* perlu diwaspadai lebih lanjut karena memiliki resiko yang lebih tinggi untuk melakukan tindakan kriminal kembali.

# Daftar pustaka

[*https://lifelines.readthedocs.io/en/latest/lifelines.datasets.html?highlight=datasets*](https://lifelines.readthedocs.io/en/latest/lifelines.datasets.html?highlight=datasets) *https://lifelines.readthedocs.io/en/latest/index.html*