





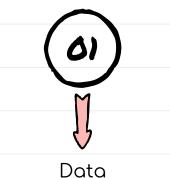


**RADR** 

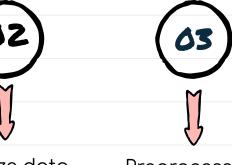












Preprocessing data



Develop Classification Model



Give Recommendation





### LATAR BELAKANG

Industri garmen merupakan industri padat karya dengan banyak proses manual. Pemenuhan demand global untuk produk garmen bergantung pada produksi dan performa karyawan dalam industri garmen.

Dataset industri garmen ini dapat digunakan untuk tujuan regresi dengan memprediksi range produktivitas (0-1) atau klasifikasi dengan mengubah range productivity (0-1) menjadi kelas-kelas berbeda.

**Tujuan project**: memprediksi aktivitas mana yang productive (1) dan non-productive (0).

## DATA UNDERSTANDING

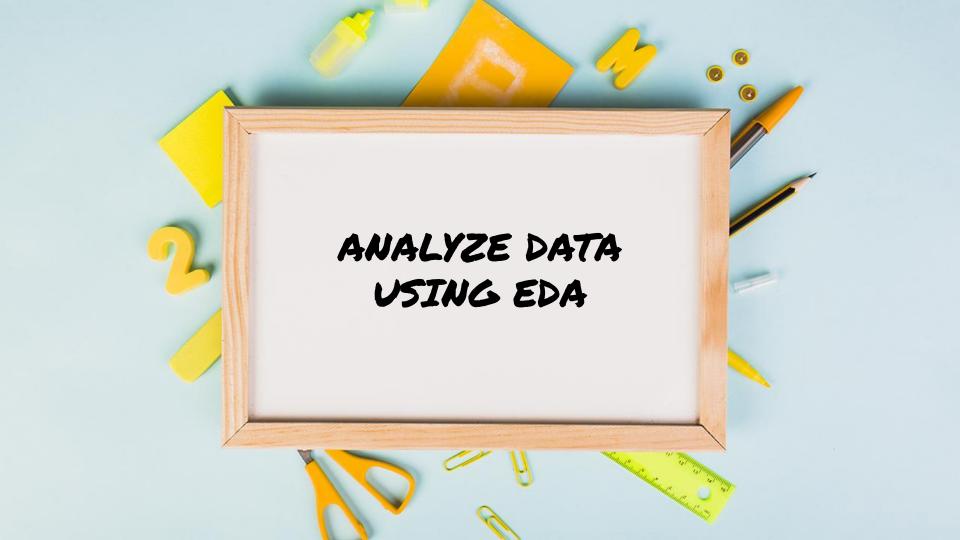


No	Atribut	Definisi
1	date	Tanggal dalam MM-DD-YY
2	day	Hari dalam satu minggu
3	quarter	Quarter merupakan bagian dari bulan dimana sebulan dibagi menjadi 4 bagian
4	department	Nama departemen terkait
5	team	Nomor tim terkait
6	no_of_workers	Jumlah pekerja setiap tim pada periode tertentu
7	no_of_style_change	Jumlah perubahan gaya (style) pada produk tertentu
8	targeted_productivity	Produktivitas yang ditargetkan oleh atasan untuk setiap tim tiap harinya.

# DATA UNDERSTANDING (CONT'D)



No	Atribut	Definisi					
9	smv	Standard Minute Value, yaitu waktu yang dialokasikan untuk suatu tugas (task)					
10	wip	Work in Progress, termasuk jumlah item yang belum selesai (unfinished items) untuk produk.					
11	over_time	Jumlah overtime setiap tim dalam satuan menit.					
12	incentive	Jumlah insentif finansial (dalam BDT) yang memungkinkan atau memotivasi tindakan tertentu.					
13	idle_time	Jumlah waktu ketika produksi diinterupsi oleh beberapa alasan.					
14	idle_men	Jumlah pekerja yang sedang tidak bekerja (idle) karena adanya interupsi dalam produksi.					
15	actual_productivity	Persentase produktivitas sesungguhnya yang diberikan pekerja. Memiliki range dari 0-1.					



## ATTRIBUTE TYPE



No	Atribut	Data Type	Tipe	Sub-tipe	Comment
1	date	object	Categorical	Ordinal	Data dimulai dari tanggal 1 Januari 2015 - 11 Maret 2015
2	day	object	Categorical	Ordinal	Terdapat 6 data yaitu Monday, Tuesday, Wednesday, Thursday, Saturday, Sunday.
3	quarter	object	Categorical	Ordinal	Terdapat 5 data yaitu Quarter 1, Quarter 2, Quarter 3, Quarter 4, Quarter 5
4	department	object	Categorical	Nominal	Terdapat 2 data yaitu Sweing dan Finishing. Namun ada data 'finishing ' yang dapat digabung bersama 'finishing'.
5	team	int64	Categorical	Ordinal	Terdapat 12 data yaitu 1-12
6	no_of_workers	float64	Numerical	Diskrit	Data bervariasi dari 2 - 89.  Ada beberapa data desimal yang muncul (misal : 51,5) sehingga ada kemungkinan error karena jumlah pekerja seharusnya bulat
7	no_of_style_change	int64	Numerical	Diskrit	Terdapat 3 data yaitu 1, 2, 3
8	targeted_productivity	float64	Numerical	Continuous	Data bervariasi dari 0,07 - 0,8

# ATTRIBUTE TYPE (CONT'D)



No	Atribut	Data Type	Tipe Data	Sub-tipe	Comment
9	smv	float64	Numerical	Continuous	Data bervariasi dari 2,9 - 54,26
					Data bervariasi dari 7 - 21.322
10	wip	float64	Numerical	Diskrit	Range terlalu besar sehingga
					kemungkinan ada outlier. Q1 - Q3 ada di
					kisaran 774,5 - 1.252,5
					Data bervariasi dari 0 - 25.920
11	over_time	int64	Numerical	Diskrit	Range terlalu besar sehingga
					kemungkinan ada outlier karena data
					terbanyak hanya di sekitar 960 dan 1.440
					Data bervariasi dari 0 - 3.600
12	incentive	int64	Numerical	Continuous	Range terlalu besar sehingga
					kemungkinan ada outlier karena data
					terbanyak adalah 0.



# ATTRIBUTE TYPE (CONT'D)

No	Atribut	Data Type	Tipe Data	Sub-Tipe	Comment
13	idle_time	float64	Numerical	Continuous	Data bervariasi dari 0 - 300  Range terlalu besar sehingga kemungkinan ada outlier karena data terbanyak adalah 0.
14	idle_men	int64	Numerical	Diskrit	Data bervariasi dari 0 - 45  Range terlalu besar sehingga kemungkinan ada outlier karena data terbanyak adalah 0.
15	actual_productivity	float64	Numerical	Continuous	Data bervariasi dari 0,23 - 1,12

### PRE-PROCESSING AWAL



#### MENGUBAH VALUE PADA KOLOM TERTENTU.

- Membulatkan ke atas (Ceiling) untuk value pada No\_of\_Workers karena ada beberapa value berupa koma.
- Mengubah value pada Department dimana 'Sweing ' diganti menjadi 'Sweing' agar tidak ada value departemen yang dobel.
- Mengganti typo untuk kata 'Sweing' menjadi 'Sewing'.

#### MISSING VALUES TREATMENT

 Mengganti NaN pada kolom WIP dengan median karena WIP tidak terdistribusi normal dan merupakan data numerik.

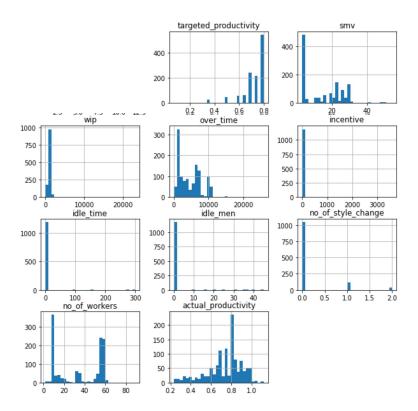
#### MENCARI DATA DUPLIKASI

Tidak ada data duplicate setelah dicek.



# ALL DISTRIBUTIONS (NUMERICAL ATTRIBUTE)

- Hampir seluruh atribut memiliki skewness tinggi baik positif maupun negatif, kecuali actual\_productivity yang terlihat memiliki distribusi normal
- Idle time dan idle workers mayoritas 0 di setiap waktu.
- Incentive yang diberikan kepada pekerja mayoritas 0 di setiap waktu.





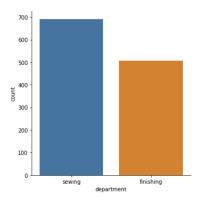
# ALL DISTRIBUTIONS NUMERICAL ATTRIBUTE (CONT'D)

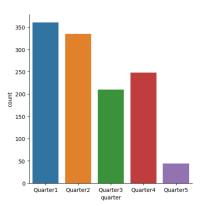
- Perubahan style (no\_of\_style\_change) hampir jarang dilakukan (0-2 kali).
- Jumlah pekerja (no\_of\_workers) bervariasi, mulai dari 8 hingga 58 pekerja.
- targeted\_productivity ditargetkan 0.8 hampir setiap waktu.
- Alokasi waktu untuk suatu tugas (smv) memiliki median 15,26 menit, sementara mayoritas tugas memiliki waktu sekitar 3 menit.

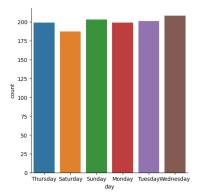


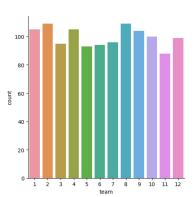
## BAR CHART (CATEGORICAL ATTRIBUTE)

- Aktivitas produksi tersebar merata sepanjang hari (day) kecuali hari Jumat (kemungkinan libur).
- Total aktivitas dept. sewing lebih banyak daripada finishing dengan selisih 200.
- Lebih dari setengah aktivitas produksi terjadi selama 2 kuarter pertama per bulannya.











# APAKAH ACTUAL PRODUCTIVITY TERDISTRIBUSI NORMAL?

Skewness : -0.807492 Kurtosis : 0.333227

Maka, distribusi dari **actual\_productivity** selaku target variable tidak normal karena *skewness* < 0 (negative skewness) dan kurtosis kurang dari 3 (platykurtic)

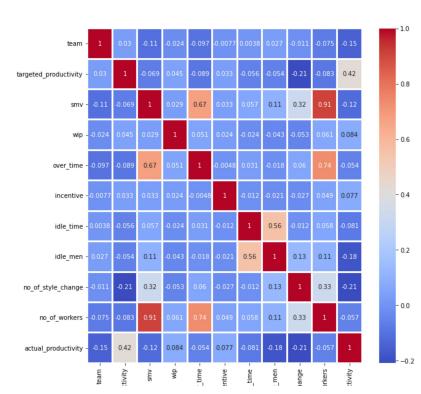
Ke depannya dapat dilakukan scaling normalization.



### CORRELATION MAP

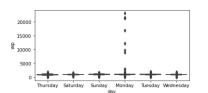
Hampir semua korelasi antar atribut lemah dan tidak signifikan. Hanya ada sedikit korelasi positif yang kuat atau moderat:

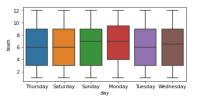
- smv dan no\_of\_workers punya korelasi positif yang kuat (0.91). Semakin lama waktu yang dibutuhkan untuk mengerjakan suatu tugas, makin banyak pekerja yang mengerjakan tugas tersebut
- **overtime** dan **no\_of\_workers** punya korelasi positif yang moderat (0.73), begitu juga **overtime** dan **smv** (0.67).
- idle\_time dan idle\_men punya korelasi positif moderat (0.56), kemungkinan karena mayoritas value bernilai 0.

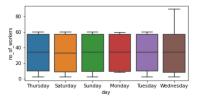


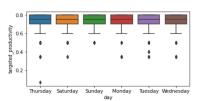
# HUBUNGAN DAY TERHADAP ATRIBUT NUMERIK

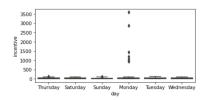
- Tidak ada perbedaan hari tiap minggunya untuk atribut tim.
- Tidak ada perbedaan hari tiap minggunya untuk jumlah pekerja. Namun, jumlah pekerja paling banyak pada hari Rabu.
- Senin memiliki outlier WIP lebih banyak daripada hari lainnya, kemungkinan karena Senin adalah awal mula periode kerja.
- Senin memiliki outlier incentive daripada hari lainnya.
- Senin dan Selasa memiliki actual productivity lebih tinggi dibanding hari lainnya, namun masih banyak outlier pada kedua hari tersebut.

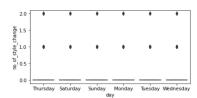


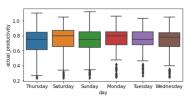


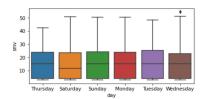






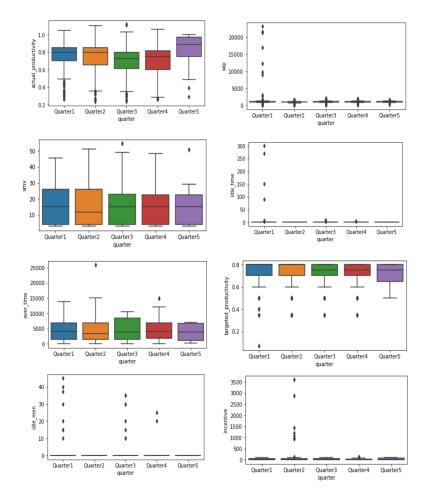






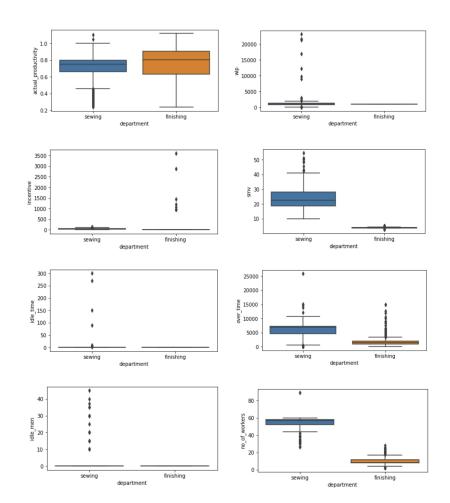
## HUBUNGAN QUARTER TERHADAP ATRIBUT NUMERIK

- Quarter 5 punya actual\_productivity lebih tinggi dibandingkan quarter lainnya.
- Quarter 1 punya outlier paling banyak pada atribut wip, idle\_men, dan idle\_time.
   Kemungkinan karena quarter tersebut adalah periode awal produksi.
- Quarter 2 punya outlier paling banyak untuk incentive.
- Quarter 3 dan 4 punya actual\_productivity lebih rendah dibandingkan quarter lainnya.
   Selain itu, outlier pada idle\_men lebih banyak terjadi pada quarter tersebut.



## HUBUNGAN DEPARTMENT TERHADAP ATRIBUT NUMERIK

- SMV lebih lama terjadi pada Dept. Sewing department daripada Dept. Finishing.
- Dept. Sewing punya outlier lebih banyak pada atribut wip, idle\_time, dan idle\_men.
- Dept. Sewing rata-rata bekerja overtime lebih lama.
- Dept. Sewing punya jumlah pekerja lebih banyak (avg = 60) daripada Dept. Finishing (avg = 12).
- Dept. Finishing punya lebih banyak outlier pada incentive
- Secara rata-rata, kedua departemen setara dalam hal actual\_productivity, namun range actual\_productivity untuk Dept. Sewing lebih konsisten dengan banyak outlier yang kurang produktif.







### FEATURE ENGINEERING

Terdapat 2 atribut baru yang dibuat yaitu atribut Performance dan Productivity.

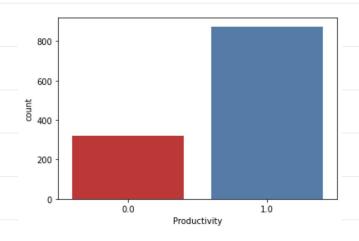
#### Performance =

Actual\_Productivity - Targeted\_Productivity

### Productivity =

Performance yang dikategorikan dalam class

- Jika Performance < 0, berarti performance bersifat non-productive (0).
- Jika Performance >= 0, berarti
   performance bersifat productive (1).



Terlihat bahwa terdapat **800** aktivitas yang tergolong **productive**, namun masih ada sekitar **300** aktivitas yang masih tergolong **non-productive**.



# CABEL ENCODING DAN

Label Encoding dilakukan ke variabel 'department' karena variabel hanya terdiri dari 2 value (sewing dan finishing).

One Hot Encoding dilakukan ke variabel 'quarter' dan 'day' karena variabel-variabel tersebut memiliki lebih dari 2 value dan tidak terpengaruh oleh order untuk dilakukan pembobotan.

Ī,	department	team	targeted_productivity	smv	wip	over_time	incentive	idle_time
Г	1	8	0.80	26.16	1108.0	7080	98	0.0
ı	0	1	0.75	3.94	1039.0	960	0	0.0
L	1	11	0.80	11.41	968.0	3660	50	0.0
ı	1	12	0.80	11.41	968.0	3660	50	0.0
ı	1	6	0.80	25.90	1170.0	1920	50	0.0
I								
ı	0	10	0.75	2.90	1039.0	960	0	0.0
ı	0	8	0.70	3.90	1039.0	960	0	0.0
t	0	7	0.65	3.90	1039.0	960	0	0.0
	0	9	0.75	2.90	1039.0	1800	0	0.0
	0	6	0.70	2.90	1039.0	720	0	0.0
•								

idle_men	 Quarter2	Quarter3	Quarter4	Quarter5	Monday	Saturday	Sunday	Thursday
0	 0	0	0	0	0	0	0	1
0	 0	0	0	0	0	0	0	1
0	 0	0	0	0	0	0	0	1
0	 0	0	0	0	0	0	0	1
0	 0	0	0	0	0	0	0	1
0	 1	0	0	0	0	0	0	0
0	 1	0	0	0	0	0	0	0
0	 1	0	0	0	0	0	0	0
0	 1	0	0	0	0	0	0	0
0	 1	0	0	0	0	0	0	0



## TRAIN TEST SPLIT

1	ar	: ~	h ~	ı	V	١
V	СI	ıu	oe	ι	$\wedge$	

- department
  - IC
- no\_of\_style\_change
- Saturday

- smv

- no\_of\_workers - Quarter1 Sunday Thursday

- WIP

- Quarter2

- Tuesday

- over\_time

team

- Quarter3

- Wednesday

- incentive

- Quarter4

- idle\_time

- Quarter5

- idle\_men
- Monday

### Variabel Y:

Productivity

X\_train.shape

(957, 21)

Y\_train.shape

(957,)

X\_test.shape

(240, 21)

Y\_test.shape

(240,)



### **SCALING**

Scaling dilakukan menggunakan normalization karena distribusi seluruh atribut tidak normal

department	team	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change
1.0	0.363636	0.363674	0.060437	0.402778	0.013889	0.0	0.0	0.0
1.0	0.000000	0.407731	0.063941	0.266204	0.031389	0.0	0.0	0.0
1.0	0.454545	0.330216	0.021804	0.069444	0.006389	0.0	0.0	0.0
1.0	0.000000	0.407731	0.066580	0.000000	0.024444	0.0	0.0	0.0
0.0	0.636364	0.000000	0.044646	0.062500	0.000000	0.0	0.0	0.0
1.0	0.636364	0.550707	0.022799	0.263889	0.015556	0.0	0.0	0.0
1.0	0.181818	0.565254	0.054250	0.268519	0.008333	0.0	0.0	0.5
1.0	0.727273	0.330216	0.068354	0.175926	0.008333	0.0	0.0	0.5
1.0	0.545455	0.443890	0.060264	0.259259	0.000000	0.0	0.0	0.0
1.0	0.909091	0.148587	0.004153	0.000000	0.012500	0.0	0.0	0.0





## KNN

Menggunakan parameter n\_neighbors = 3

Evaluation	Mo	odel: k	(NN Trainin	ng	Evaluation	Mo	odel:	KNN	Testing	
[[153 105]					[[ 20 44]					
[ 39 660]]					[ 31 145]	]				
Accuracy	:	84.95	%		Accuracy	:	68.75	%		
Precision	:	86.27	%		Precision	:	76.72	%		
Recall	:	94.42	%		Recall	:	82.39	%		



## DECISION TREE

Menggunakan parameter max\_depth = 8

```
Evaluation Model: Decision Tree Training Evaluation Model: Decision Tree Testing
[[199 59] [[ 38 26] [ 27 149]]
Accuracy : 90.8 %
```

Accuracy : 90.8 % Accuracy : 77.92 % Precision : 91.91 % Precision : 85.14 %

Recall : 95.85 % Recall : 84.66 %



## RANDOM FOREST

Recall : 98.86 %

Menggunakan parameter max\_depth = 8

Recall : 92.61 %

### Random Forest punya model paling baik.

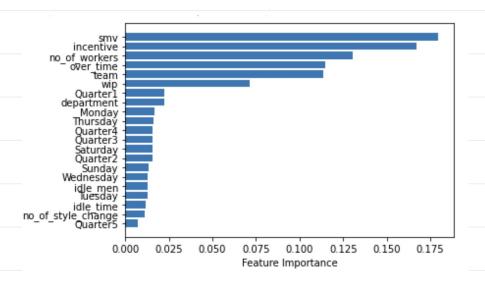
Selanjutnya model ini dapat diberikan treatment SMOTEENN (karena model memiliki imbalanced data) dan Hyperparameter Tuning



### FEATURE IMPORTANCE

Dari model Random Forest, diketahui bahwa 3 atribut paling berpengaruh dalam pembuatan model adalah:

- SMV
- Incentive
- No of Workers





# RANDOM FOREST WITH SMOTEENN

 Evaluation Model: Random Forest Training
 Evaluation Model: Random Forest testing

 [[459 0]
 [[50 14]

 [ 0 370]]
 [ 52 124]]

 Accuracy : 100.0 %
 Accuracy : 72.5 %

 Precision : 100.0 %
 Precision : 89.86 %

 Recall : 100.0 %
 Recall : 70.45 %

Ternyata setelah model Random Forest diberikan treatment SMOTEENN, parameter accuracy dan recall dari testing model mengalami penurunan jika dibandingkan dengan model tanpa SMOTEENN.

# CLASSIFICATION REPORT WITHOUT SMOTEENN

## CLASSIFICATION REPORT WITH SMOTEENN

	precision	recall	f1-score	support		precision	recall	f1-score	support
0.0 1.0	0.69 0.85	0.56 0.91	0.62 0.88	64 176	0.0		0.78 0.70	0.60 0.79	64 176
accuracy macro avg weighted avg	0.77 0.81	0.74 0.82	0.82 0.75 0.81	240 240 240	accuracy macro avg weighted avg	0.69	0.74 0.72	0.73 0.70 0.74	240 240 240

### Recall dari 0.0 (non-productive) mengalami peningkatan dari 0.56 menjadi 0.78

Artinya, kemungkinan pekerja non-productive disalahartikan sebagai productive berkurang. Hal ini baik karena perusahaan memerlukan model yang dapat mendeteksi pekerja non-productive sebanyak mungkin.

Meskipun begitu, banyak parameter lainnya menunjukkan penurunan sehingga model yang dipilih adalah model tanpa treatment SMOTEENN.



# HYPERPARAMETER TUNING

Selanjutnya, model Random Forest tanpa SMOTEENN diberikan treatment Hyperparameter Tuning.

Model terbaik menggunakan parameter max\_features = 8 dan n\_estimators = 110.

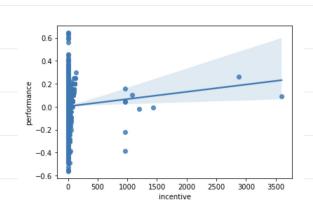
F1-Score menunjukkan peningkatan dari 0,75 menjadi 0,77.



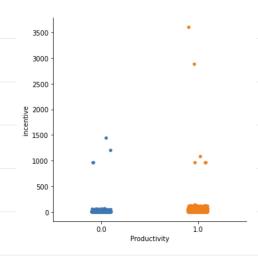




# REKOMENDASI I : MENAMBAH INCENTIVE



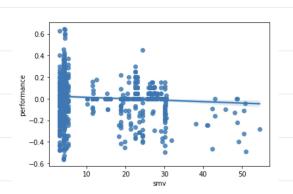
Berdasarkan regplot di atas, terdapat tren naik pada incentive terhadap Performance. Oleh karena itu, sebaiknya incentive ditambah agar produktivitas dapat ditingkatkan.

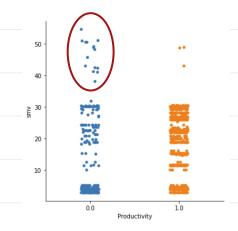


Berdasarkan catplot di atas, aktivitas productive lebih banyak terjadi saat\_incentive lebih dari 0.



# REKOMENDASI II : MENGURANGI SMV

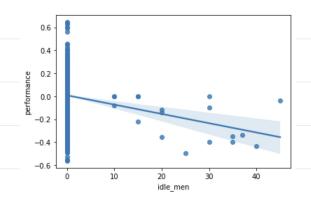


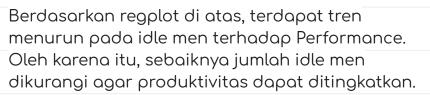


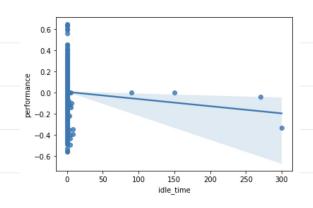
Berdasarkan regplot di atas, terdapat tren menurun pada SMV terhadap Performance. Oleh karena itu, sebaiknya SMV dikurangi agar produktivitas dapat ditingkatkan. Berdasarkan catplot di atas, aktivitas non-productive lebih banyak terjadi saat SMV berada di sekitar 40 - 50.



# REKOMENDASI III : MENGURANGI IDLE MEN DAN IDLE TIME







Berdasarkan regplot di atas, terdapat tren menurun pada idle time terhadap Performance. Oleh karena itu, sebaiknya idle time dikurangi agar produktivitas dapat ditingkatkan.



### POTENSI PENELITIAN LANJUTAN

Selanjutnya, perusahaan dapat melakukan penelitian mengenai :

- Batasan incentive yang dibutuhkan untuk meningkatkan produktivitas maksimal.
- Batasan waktu yang perlu dikurangi dalam
   SMV untuk meningkatkan produktivitas
   secara maksimal.



