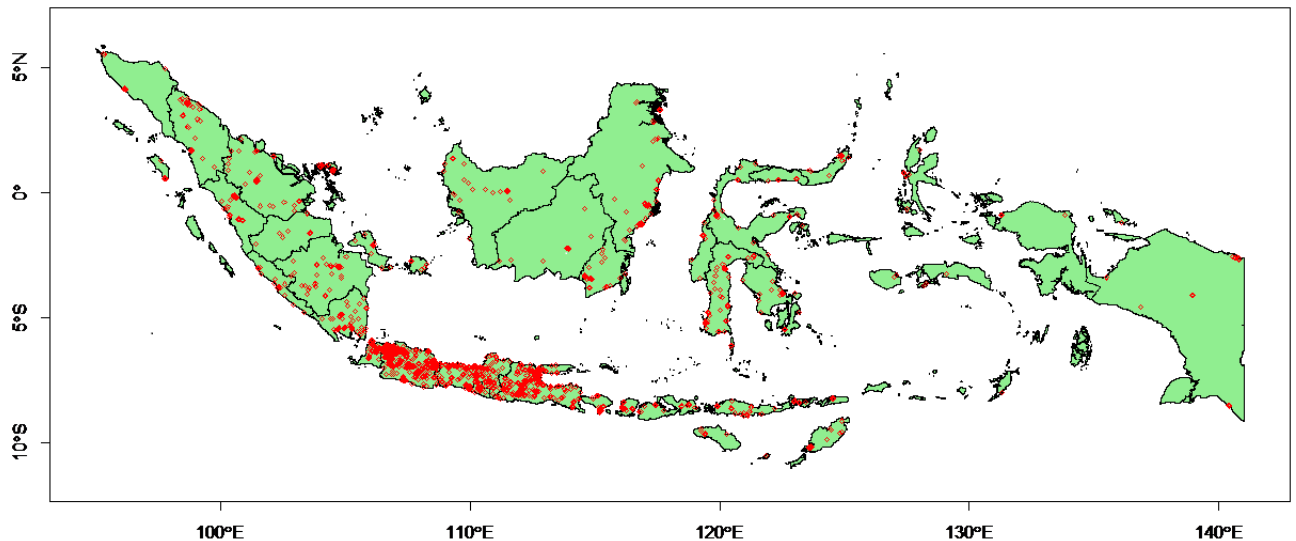


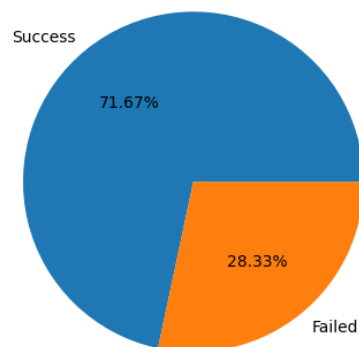
Insights of Task Delivery Data

1. Jika dilakukan pemetaan pada wilayah Indonesia, tugas pengiriman paling banyak terdapat pada Pulau Jawa, sedangkan tugas pengiriman paling sedikit terdapat pada Pulau Papua (jika didasarkan pada 5 pulau terbesar di Indonesia).



2. Dengan tidak melibatkan tugas yang masih dalam proses (*ongoing*), dapat diketahui bahwa proporsi antara tugas yang sukses dengan tugas yang gagal cukup jauh berbeda. Jika terdapat 100 tugas pengiriman yang dikerjakan, kemungkinan akan terdapat sekitar 29 tugas pengiriman yang gagal.

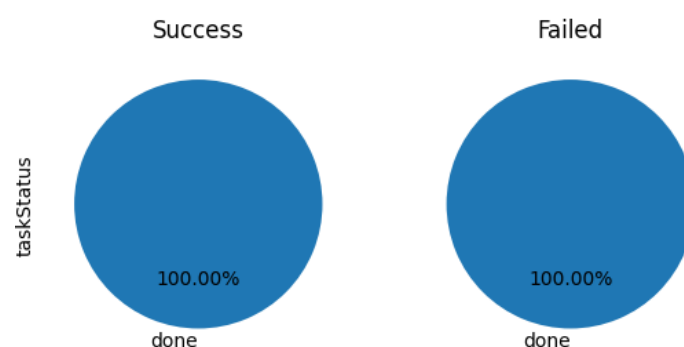
Pie Chart : Task Status Label



Namun, karena pengguna dari jasa pengiriman jauh lebih banyak dari 100 pengguna, maka tentu saja kondisi ini kurang ideal untuk terus dapat mempertahankan pengguna jasa pengiriman. Sehingga nantinya diperlukan upaya untuk meminimalisir terjadinya kegagalan dalam pengerjaan tugas pengiriman. Salah satu upaya yang dapat dilakukan adalah dengan memahami faktor/variabel apa saja yang sangat berpengaruh terhadap keberhasilan tugas pengiriman. Selain itu, dapat juga dilakukan tindakan preventif dengan memprediksi suatu tugas apakah akan sukses atau gagal, lalu pada tugas-

tugas yang diprediksi gagal dapat diberikan perhatian atau *treatment* tambahan agar prediksi kegagalan tersebut tidak terjadi di lapangan.

3. Semua tugas pengiriman dapat dikonfirmasi sukses atau gagal hanya jika tugas pengiriman tersebut telah dikonfirmasi selesai. Jika tugas pengiriman masih dalam proses (*ongoing*), maka belum dapat dikonfirmasi apakah tugas tersebut sukses atau gagal. Selain itu, kedua proporsi variabel *task status* antara kategori tugas yang sukses dan gagal adalah sama. Dengan demikian, jika tujuan analisis adalah untuk melakukan klasifikasi, variabel ini tidak dapat dijadikan sebagai *classifier variable*. Hal ini dikarenakan variabel *task status* sama sekali tidak berperan dalam membedakan tugas mana yang berhasil atau gagal.



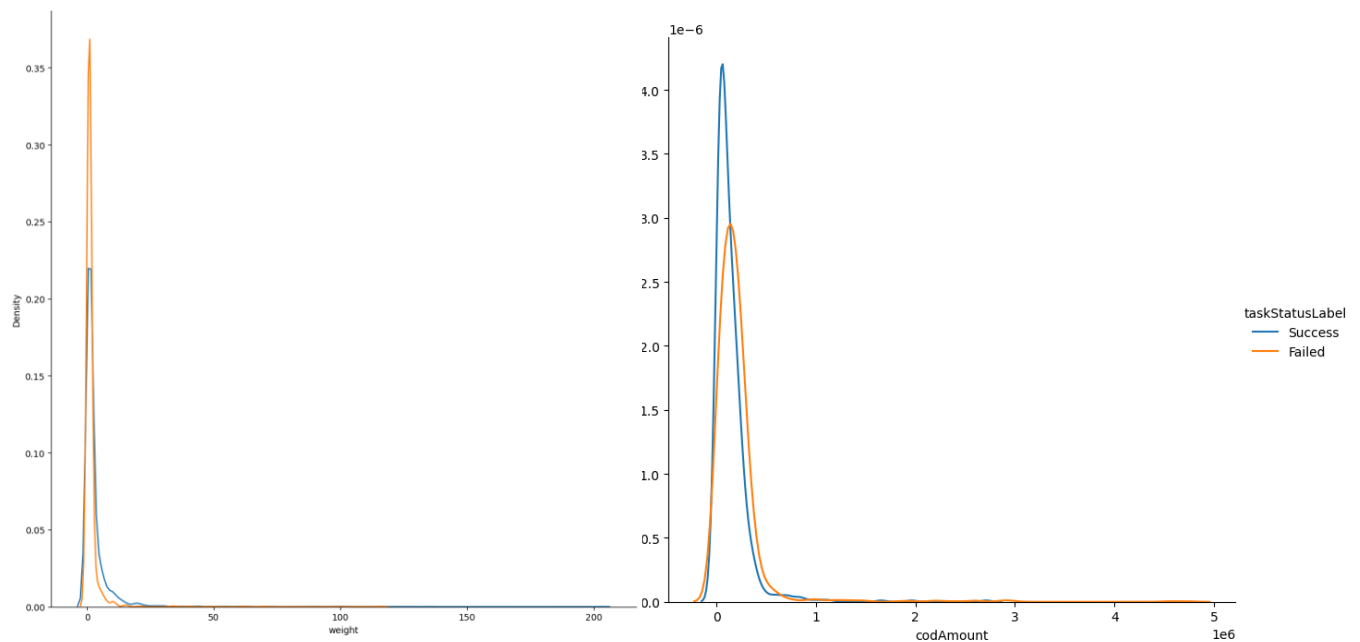
4. Variabel *cod amount* dan *weight* memiliki korelasi/hubungan yang positif. Hal ini menunjukkan bahwa jika berat barang yang dikirim bertambah, maka *cod amount* yang harus ditanggung juga akan semakin besar. Namun, dengan nilai korelasi sebesar 0.44, hubungan positif antara kedua variabel tersebut dapat dikatakan tidak terlalu kuat ataupun tidak terlalu lemah (medium). Hal ini memungkinkan untuk terjadi. Walaupun sering ditemukan kecenderungan bahwa semakin berat suatu barang yang dikirim/dibeli maka akan semakin besar pula harga yang harus dibayar. Namun terdapat barang-barang tertentu yang berbeda dengan kecenderungan yang ada, seperti misalnya emas yang disandingkan dengan barang lain yang memiliki berat yang sama.

	<i>codAmount</i>	<i>weight</i>
<i>codAmount</i>	1.000000	0.439865
<i>weight</i>	0.439865	1.000000

Dalam melakukan analisis lebih lanjut dimana variabel *cod amount* dan *weight* dipertimbangkan untuk digunakan, karena nilai korelasi yang tidak terlalu kuat terjadi antara kedua variabel, maka kedua variabel ini tidak dapat saling menggantikan satu sama lain.

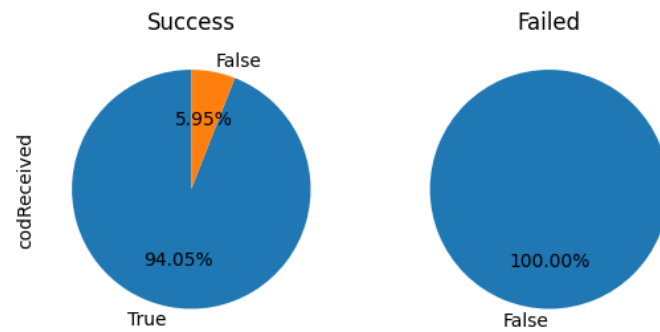
5. Jika akan dilakukan pembuatan model klasifikasi, secara visual variabel *cod amount* dan *weight* berpotensi untuk dihilangkan. Hal ini dapat terlihat dari kurva distribusi masing-masing variabel pada

setiap kategori sukses dan gagal. Kurva dari masing-masing variabel berhimpitan dan memiliki letak puncak yang relatif sama (*measures of central tendency* seperti median mirip). Sehingga secara visual, kedua variabel dapat dianggap tidak terlalu berperan dalam membedakan kategori tugas yang sukses dan gagal.

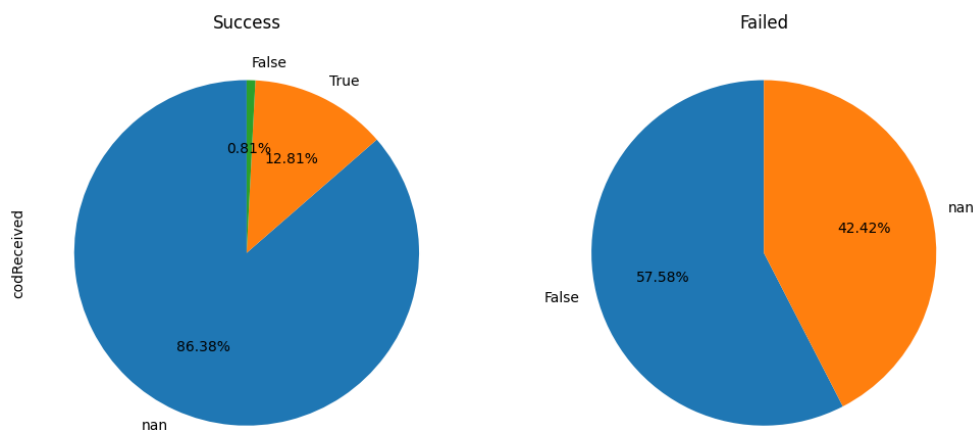


Namun, diperlukan pengujian menggunakan metode statistika yakni menggunakan *t-student test* agar kesimpulan yang diperoleh merupakan hasil analisis secara objektif. Hasil dari *t-student test* yang membandingkan antara kelompok *cod amount* yang sukses mengerjakan tugas dengan kelompok data *cod amount* yang gagal mengerjakan tugas, ternyata memberikan kesimpulan bahwa kedua kelompok tersebut berbeda. Kesimpulan yang sama juga diperoleh setelah membandingkan antara kelompok *weight* yang sukses mengerjakan tugas dengan kelompok *weight* yang gagal mengerjakan tugas. Dengan demikian, variabel *cod amount* dan *weight* dapat digunakan dalam analisis, khususnya klasifikasi.

6. Jika hanya melibatkan tugas pengiriman yang menggunakan sistem pembayaran COD, semua tugas yang dianggap gagal merupakan tugas yang dikonfirmasi bahwa COD tidak diterima. Sedangkan pada tugas pengiriman yang dianggap berhasil, memang kebanyakan telah dikonfirmasi bahwa COD telah diterima, namun ada sekitar 5,95% dari COD yang dikonfirmasi tidak diterima namun tetap dilaporkan bahwa tugas pengiriman telah berhasil. Hal ini tentunya merupakan sebuah kondisi yang janggal dan memerlukan peninjauan lebih lanjut terkait alasan COD tidak diterima pengguna namun dilaporkan berhasil.



Jika tidak hanya melibatkan semua tugas pengiriman yang tertera menggunakan sistem COD, maka dapat terlihat pada gambar di bawah bahwa terdapat banyak sekali *missing value* dalam variabel *cod received*. *Missing value* ini dapat terjadi karena berbagai macam alasan seperti misalnya pengguna tidak menggunakan sistem COD dalam pembayarannya atau pengguna menggunakan sistem pembayaran COD namun lupa atau tidak mengkonfirmasi pesanan yang telah diterima, atau memang nilai tersebut benar-benar tidak ada. Karena *missing value* yang terdapat pada variabel *cod received* sangat banyak, maka kemungkinan besar variabel ini tidak akan dimasukkan dalam analisis seperti misalnya pembuatan model klasifikasi.

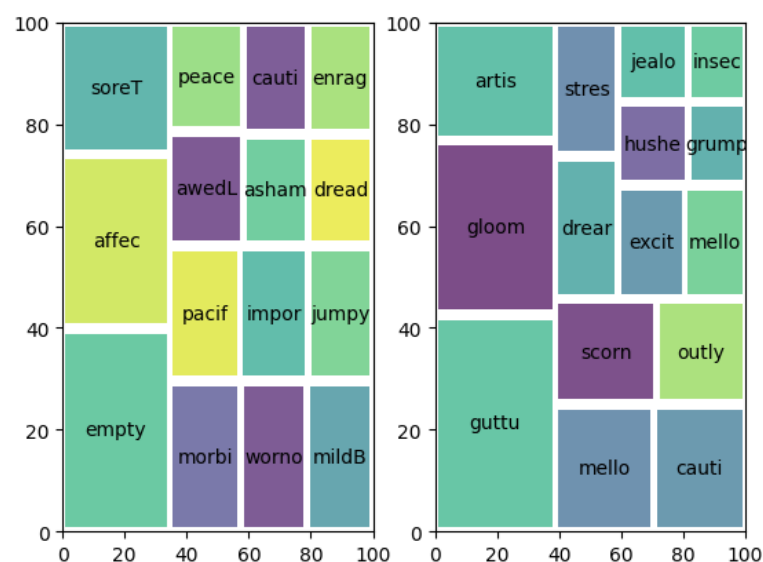


7. Pekerja yang paling banyak berhasil menyelesaikan tugas pengiriman dan layak direkomendasikan untuk diberi kepercayaan lebih dalam mengerjakan tugas, serta pekerja yang paling banyak gagal dalam menyelesaikan tugas pengiriman dan sangat tidak disarankan untuk diberikan tugas pengiriman di kemudian hari adalah:

	Worker	Task Failed
0	gutturalLion9	103
1	gloomyLlama0	83
2	artisticHyena7	56
3	mellowDoughnut1	49
4	cautiousUnicorn6	46
5	scornfulOcelot0	42
6	outlyingDoughnut6	37
7	drearyPepper9	35
8	stressedRat6	33
9	excitedBustard4	30

	Worker	Task Success
0	emptyIcecream6	57
1	affectedKitten8	49
2	soreTomatoe9	37
3	morbidFlamingo7	28
4	wornoutEggs4	26
5	mildBoa2	26
6	pacifiedLion0	25
7	importedStork6	24
8	jumpyCockatoo6	23
9	awedLemur8	22

Jika merujuk pada gambar *heatmap* di bawah, terlihat bahwa pekerja yang paling banyak berhasil dan pekerja yang paling banyak gagal, masing-masing terdiri dari identitas pekerja yang berbeda satu sama lain. Kesimpulan ini juga telah didukung oleh pengecekan langsung terhadap semua nilai pada variabel *task assigned to*. Sehingga, variabel *task assign to* cocok digunakan sebagai variabel yang dapat membedakan berhasil atau tidaknya suatu tugas pengiriman. Keputusan ini didukung pula oleh pengujian menggunakan metode statistika *chi-squared test* yang membuktikan bahwa kelompok data pekerja yang berhasil dan kelompok data pekerja yang gagal berbeda secara signifikan.

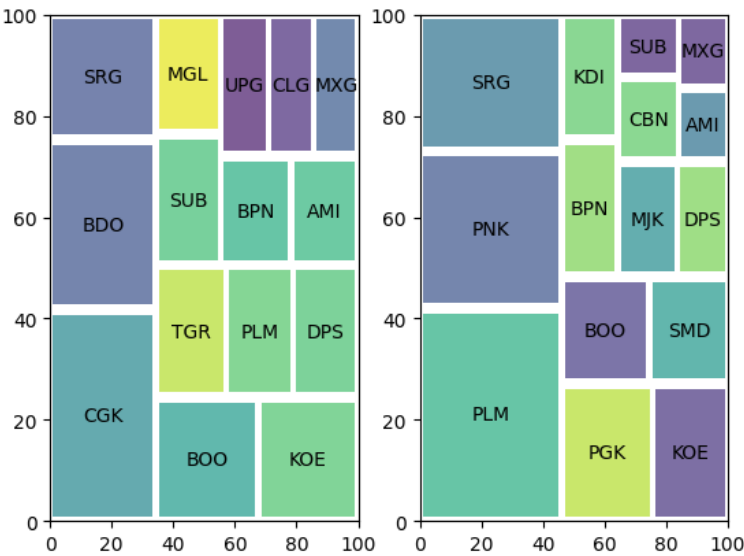


8. Terdapat 3 cabang destinasi yang tidak pernah memiliki catatan kegagalan dalam tugas pengiriman. 3 cabang destinasi tersebut adalah PBL, SDA, dan BTG.

Cabang destinasi yang memiliki catatan keberhasilan tugas pengiriman terbanyak, serta cabang destinasi yang memiliki catatan keberhasilan tugas pengiriman paling sedikit (cabang destinasi yang perlu ditinjau lebih lanjut terkait penyebab terjadinya kegagalan tugas pengiriman) adalah:

Branch Destination Task Success			Branch Destination Task Failed		
0	CGK	476	0	PLM	307
1	BDO	383	1	PNK	226
2	SRG	282	2	SRG	197
3	BOO	270	3	PGK	127
4	KOE	263	4	KOE	105
5	TGR	199	5	BOO	95
6	PLM	188	6	SMD	85
7	DPS	183	7	BPN	78
8	SUB	182	8	KDI	71
9	MGL	166	9	MJK	70

Jika merujuk pada gambar *heatmap* di bawah yang memperlihatkan komposisi antara cabang destinasi yang terekam paling banyak berhasil dan yang terekam paling banyak gagal , terdapat 9 cabang yang ada pada kedua gambar *heatmap* tersebut. 9 cabang tersebut adalah SRG, BOO, KOE, PLM, DPS, SUB, BPN, AMI, dan MXG. Hal tersebut menunjukkan bahwa kesembilan cabang yang telah disebutkan merupakan cabang destinasi yang paling banyak berhasil namun paling banyak tidak berhasil pula. Kondisi ini mungkin saja terjadi jika memang frekuensi tugas yang diberikan pada 9 cabang tersebut jauh lebih banyak jika dibandingkan cabang destinasi lainnya.



Walaupun terdapat cabang destinasi yang sama antara kedua *heatmap*, namun jika diperhatikan lebih detail, cabang-cabang destinasi pada setiap *heatmap* tersebut memiliki komposisi yang cukup

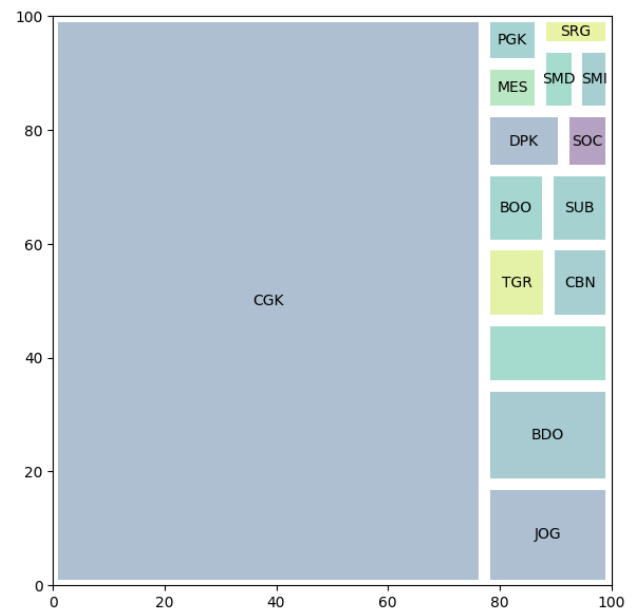
berbeda. Sehingga, variabel *branch destination* cocok digunakan sebagai variabel yang dapat membedakan berhasil atau tidaknya suatu tugas pengiriman. Keputusan ini didukung pula oleh pengujian menggunakan metode statistika *chi-squared test* yang membuktikan bahwa kelompok data cabang destinasi yang berhasil dan kelompok data cabang destinasi yang gagal berbeda secara signifikan.

9. Terdapat 3 cabang awal yang tidak pernah memiliki catatan kegagalan dalam tugas pengiriman. Ketiga cabang tersebut adalah PSR, CXP, MDC, DJJ, TTE, DJB, BTG, BTJ, BKS, SOQ, AMQ, DTB, dan PKY.

Cabang awal yang memiliki catatan keberhasilan tugas pengiriman terbanyak, serta cabang awal yang memiliki catatan kegagalan tugas pengiriman terbanyak adalah :

Branch Origin Task Success			Branch Origin Task Failed		
0	CGK	3468	0	CGK	1563
1	BDO	230	1	JOG	82
2	None	226	2	BDO	79
3	TGR	177	3	None	53
4	SUB	128	4	TGR	31
5	BOO	111	5	CBN	30
6	JOG	83	6	BOO	30
7	SRG	72	7	SUB	30
8	UPG	65	8	DPK	30
9	BKI	60	9	SOC	18

Jika merujuk pada gambar *heatmap* di bawah yang memperlihatkan komposisi antara cabang awal yang terekam paling banyak berhasil dan yang terekam paling banyak gagal , terdapat 12 cabang yang ada pada kedua gambar *heatmap* tersebut. 12 cabang tersebut adalah CGK, BDO, None, TGR, SUB, BOO, JOG, SRG, MES, SOC, DPK, dan CBN. Hal tersebut menunjukkan bahwa 12 cabang yang telah disebutkan merupakan cabang awal yang paling banyak berhasil namun paling banyak tidak berhasil pula. Kondisi ini mungkin saja terjadi jika memang frekuensi tugas yang diberikan pada 12 cabang tersebut jauh lebih banyak jika dibandingkan cabang awal lainnya.



Walaupun terdapat cabang yang sama antara kedua *heatmap*, terlihat bahwa banyak komposisi pada cabang berbeda cukup jauh antara 2 *heatmap* tersebut. Sehingga, variabel *branch origin* cocok digunakan sebagai variabel yang dapat membedakan berhasil atau tidaknya suatu tugas pengiriman. Keputusan ini didukung juga oleh pengujian menggunakan metode statistika *chi-squared test* yang membuktikan bahwa kelompok data cabang awal yang berhasil dan kelompok data cabang awal yang gagal berbeda secara signifikan.

10. Kota penerima yang memiliki catatan keberhasilan tugas pengiriman terbanyak, serta kota penerima yang memiliki catatan kegagalan tugas pengiriman terbanyak adalah :

Receiver City Task Success			Receiver City Task Failed		
0	Cidaun	75	0	Seberang ulu i	77
1	Denpasar selatan	66	1	Sungai raya	65
2	Tebet	49	2	Pontianak kota	58
3	Pituruh	47	3	Pontianak tenggara	48
4	Mataram	39	4	Tanah grogot	46
5	Sintang	36	5	Tenggarong	43
6	NaN	36	6	Plaju	36
7	Grogol	36	7	Pangkal balam	33
8	Jombang	35	8	Seberang ulu ii	32
9	Oebobo	35	9	Kadugede	32

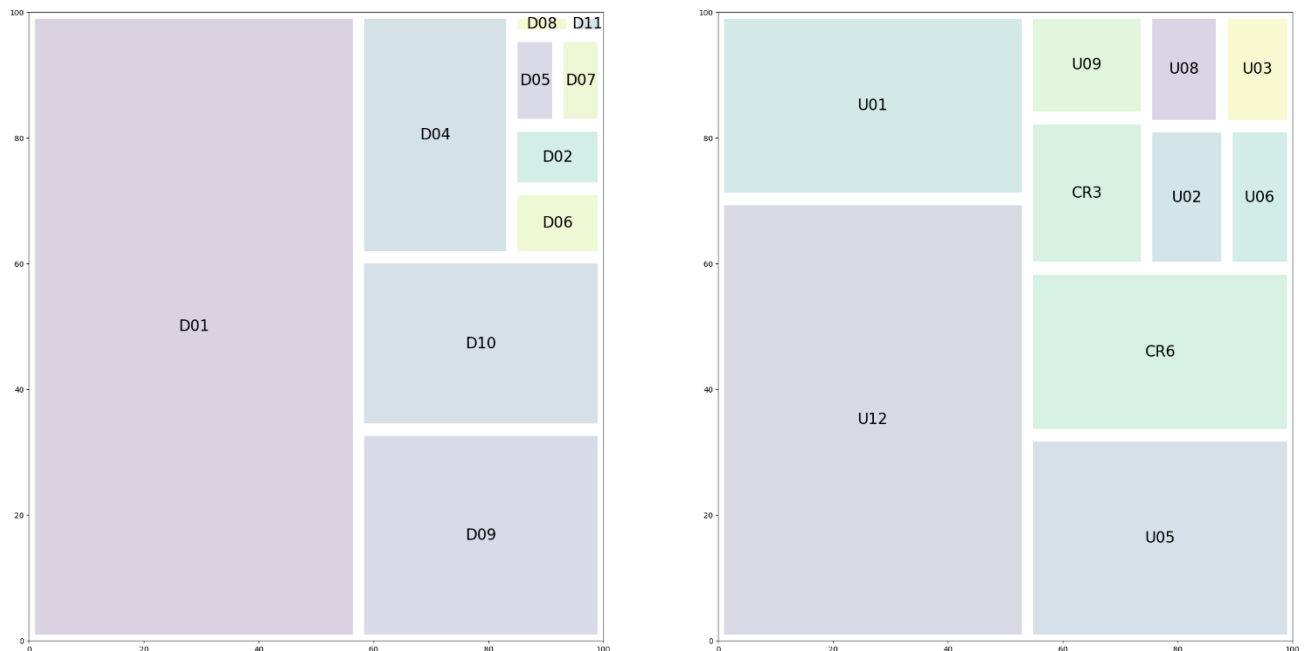
Jika merujuk pada gambar *heatmap* di bawah, terlihat bahwa kota penerima yang paling banyak berhasil dan kota yang paling banyak gagal, masing-masing terdiri dari kota yang berbeda satu sama lain. Sehingga, variabel *receiver city* cocok digunakan sebagai variabel yang dapat membedakan berhasil atau tidaknya suatu tugas pengiriman. Keputusan ini didukung juga oleh pengujian menggunakan metode statistika *chi-squared test* yang membuktikan bahwa kelompok data kota penerima yang berhasil dan kelompok data kota penerima yang gagal berbeda secara signifikan.



11. Penerima/keterangan yang paling sering ditemukan pada tugas pengiriman yang berhasil, serta penerima/keterangan yang paling sering ditemukan pada tugas pengiriman yang gagal adalah sebagai berikut:

Task Detail Status Task Success			Task Detail Status Task Failed		
0	YANG BERSANGKUTAN	3109	0	MISROUTE	763
1	KELUARGA/SAUDARA	774	1	ALAMAT TIDAK LENGKAP service/ TIDAK DIKENAL	322
2	ATASAN/STAFF/KARYAWAN/BAWAHAN	634	2	RUMAH service/ KANTOR KOSONG (MASIH DIHUNI)	304
3	SECURITY	564	3	NEW ADDRESS	247
4	SUAMI/ISTRI/ANAK	94	4	DIAMBIL SENDIRI	100
5	RECEPTIONIST	87	5	TUTUP PADA AKHIR PEKAN service/ HARI LIBUR	70
6	MAILING ROOM	62	6	PENERIMA TIDAK DIKENAL	64
7	PEMBANTU	61	7	DITOLAK OLEH PENERIMA	52
8	PENJAGA KOS	21	8	PENERIMA MENOLAK BAYAR (KIRIMAN COD)	48
9	SUPIR	11	9	PENERIMA PINDAH ALAMAT	45

Jika merujuk pada gambar *heatmap* di bawah, terlihat bahwa penerima/keterangan yang berkaitan dengan tugas paling banyak berhasil dan yang paling banyak gagal, masing-masing terdiri dari penerima/keterangan yang berbeda satu sama lain. Sehingga, variabel *task detail status* atau *task detail status label* cocok digunakan sebagai variabel yang dapat membedakan berhasil atau tidaknya suatu tugas pengiriman. Keputusan ini didukung juga oleh pengujian menggunakan metode statistika *chi-squared test* yang membuktikan bahwa kelompok data *task detail status* yang berhasil dan kelompok data *task detail status* yang gagal berbeda secara signifikan.



12. Dalam rangka melakukan antisipasi terhadap terjadinya kegagalan dalam penyelesaian tugas pengiriman, maka dibuatlah sebuah model klasifikasi yang dapat memprediksi keberhasilan tugas-tugas pengiriman yang masih dalam proses (*ongoing*). Data tugas pengiriman terlebih dahulu dipersiapkan dengan cara menghilangkan data yang terduplikasi, mengatasi *missing value*, melakukan *feature selection* dan menghapus kolom-kolom yang tidak berpengaruh terhadap variabel target, melakukan kuantifikasi pada variabel-variabel kategorik, membagi data menjadi *training data* dan *testing data*, serta melakukan standarisasi data (*scaling*). Setelah data siap digunakan, maka dilakukan pemodelan menggunakan metode *K-Nearest Neighbors* yang dikombinasikan dengan *Grid Search Cross-Validation*. Detail pembuatan model klasifikasi dapat dilihat pada link di bawah ini:

<https://colab.research.google.com/drive/1g7RbRwbOHluAPXQLFAtc5p-92g1NE-aa#scrollTo=liUaA9pDa0vF>

Performa dari model yang berhasil dibuat adalah sebagai berikut :

	Training Performance	Testing Performance
Accuracy	84.51	78.51
Precision	86.32	82.52
Recall	93.04	88.95

Karena performa model klasifikasi yang dibuat telah cukup bagus, maka model tersebut dapat digunakan untuk memprediksi tugas-tugas pengiriman yang belum selesai. Hasil prediksinya adalah sebagai berikut.

	branch_dest	taskStatusLabel	receiver_city	taskDetailStatusLabel	taskDetailStatus	weight	branch_origin	taskStatusUser	taskStatusPrediction
f5cd82456af0	MES		DOLOK PANRIBUAN,SIMA				1 CGK		Success
'e0d978e79e	MES		SIANTAR TIMUR,PEMATA				1 MES		Success
f04d5eb328e	MES		SIANTAR TIMUR,PEMATA				1 SOC		Success
fb44eb9878e	MES		SIANTAR BARAT,PEMATA				1 PDG		Success
1251136dab5	MES		SIANTAR BARAT,PEMATA				1 CGK		Success
:0327bcb95d	MES		SIANTAR SITALASARI,P				1 CGK		Success
887db9aefdc	MES		SIANTAR MARTOBA,PEMA				1 CGK		Success
:5623bf2ad6f	MES		SIANTAR,SIMALUNGUN				1 JOG		Success
3db4ca31d5c	CKR		CIKARANG BARAT,CIKAR				3 CGK		Success
738e923ea33	BKI		BEKASI UTARA, BEKASI				6 CGK		Success
1493962696e	UPG		BAJENG,SUNGUMINASA				0.2 CGK		Success
:6992e1896a	UPG		BAJENG				1 CGK		Success
19197e24907	BPN		BALIKPAPAN SELATAN,				0.25 CGK		Failed
f56de16befc6	PLM		NIBUNG,MUARA BELITI				1 CGK		Failed
0a10c224b3e	TKG		SUKADANA, KAB.LAMPUN				3 CGK		Success

Untuk hasil lebih lengkapnya dapat dilihat pada link dokumen di bawah ini:

https://docs.google.com/spreadsheets/d/1TOL9Ynt0Jr1BLx_lIOAhDMrJ4IcH874-MTCLLvQgufQ/edit?usp=sharing

Berdasarkan hasil prediksi tersebut, diketahui bahwa dari 693 tugas pengiriman yang masih dalam proses, terdapat 141 tugas pengiriman yang diprediksi akan gagal. Oleh sebab itu, perusahaan dapat melakukan tindakan-tindakan yang dapat mencegah terjadinya kegagalan pada 141 tugas pengiriman tersebut.

- Sesuai dengan hasil *permutation importance* pada model klasifikasi, untuk mencegah/meminimalisir terjadinya kegagalan pada 141 tugas pengiriman yang terprediksi gagal, perusahaan disarankan untuk meninjau ulang cabang destinasi dimana diprediksi akan gagal menyelesaikan tugas. Karena dibandingkan dengan 4 *classifier variable* lainnya, variabel cabang destinasi memiliki peran paling penting dalam keberhasilan suatu tugas pengiriman. Setelah itu, jika terdapat sumber daya yang memadai, perusahaan dapat pula melakukan peninjauan ulang pada variabel *task assign to* dan *branch origin*. Variabel lainnya yaitu *receiver city* dan *weight* merupakan variabel yang berpengaruh

terhadap keberhasilan tugas pengiriman, namun kedua variabel tersebut berada di luar kontrol perusahaan.