bank btpn A MEMBER OF SMBC Group

Income Estimation Model for NTB & Non Bureau

RAPP - Jul 2023

Background

- **Background**

  Currently Bank BTPN doesn’t have income proxy calculation for NTB and Non Bureau segment. For applicants which falls under that segment, Bank BTPN only have simple limit calculation based on expert judgement.

  In order to improve credit limit assessment method, reduce too high/low limit, and to prevent loss of business opportunity, RAPP decided to develop income estimation model.

Model Framework

Regression Formula

Y = B0 + B1 X1 + B2 X2 + B3 X3 +……

Y = Constant + B1*Age + B2*LengthofWork + ….

Y_pred	Y		Type of Predictor	is used
Prediction	Real Income (Avg 3 Months Salary)		Demography	1
			Macro Economic + External Data	1
			SDK	1

*Y_pred vs Y_actual are needed for model validation

- Model will predict the income exact value
- Model will be utilized for NTB & Non Bureau segment
- Most of predictors such as internal loan, bureau data, funding, etc are not available for NTB & Non Bureau segment

Sample Definition

- Sample from payroll account during the period of Oct 2021 – Dec 2022
- Exclusion :
  - BTPN Syariah payroll (the salary amount does not reflect the full salary payment)
  - Avg 3-month salary of < 500K or >30Mio
- Take the average from latest 3 months salary payment for each CIF

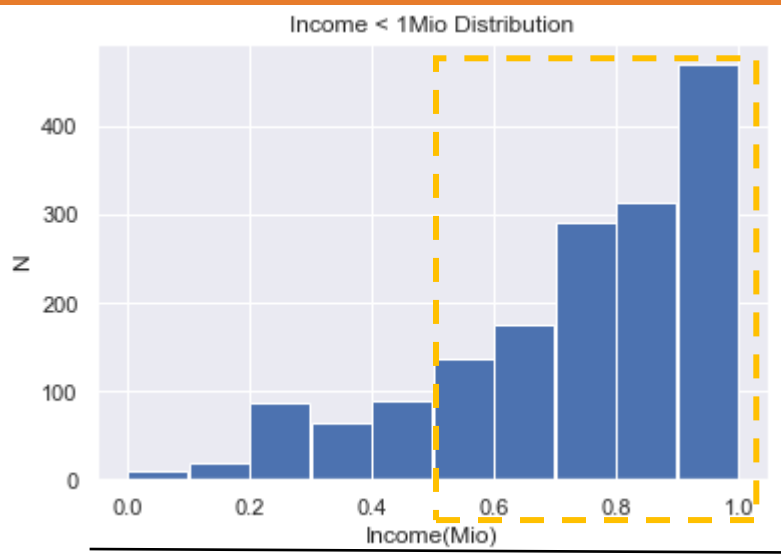
Criteria	Pop	Drop
All Payroll Oct'21 - Dec'22	302,556	
Exclude BTPN Syariah	235,536	67,020
Last 3 months for each CIF	77,726	157,810
Unique CIF	22,790	54,936
Average 3 month salary >= 500K & <=30Mio	22,204	586

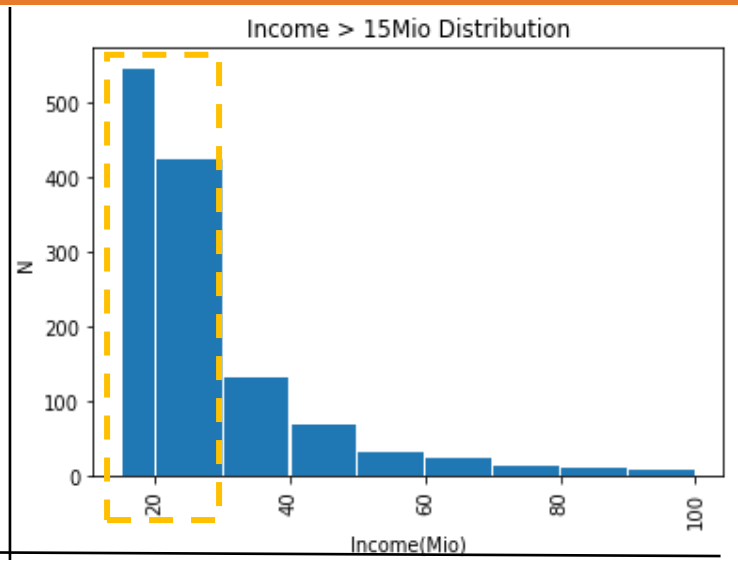
	Avg Income L3M (Mio)
Count	22,790
Mean	5.68
STD	8.94
Min	0.000001
25%	1.58
50%	3.9
75%	6.59
Max	255.12
OutlierLower	-5.93
OutlierUpper	14.11

* Model will be utilized for NTB & Non Bureau segment

Sample Definition

- 500K to 1 Mio already constitutes 84 % (1386 sample) of Income sample <1Mio
- 15Mio – 30Mio already constitutes 75% (971 sample) of Income sample > 15Mio





- Payroll sample consist of 263 companies
- If we exclude BTPN SYARIAH, there is no dominant company in the sample; hence, the payroll sample can quite represent real income population

Top 10 Company by contribution

Company_Name	NCIF	All	Excl - BTPN
PT BTPN SYARIAH	8,547	22.83%	
SENTRA TRADA INDOSTATION, PT	4,823	12.88%	16.70%
SEJAHTERA SEJATI INTIPERMATA, PT	2,060	5.50%	7.13%
INDOMOBIL FINANCE INDONESIA, PT	1,011	2.70%	3.50%
SEINO INDOMOBIL LOGISTICS, PT	967	2.58%	3.35%
PT GLOBAL GREEN TRADING	836	2.23%	2.89%
PT VALDO INTERNATIONAL	758	2.02%	2.62%
TJOA VERAWATI	733	1.96%	2.54%
YAYASAN HARAPAN BANGSA	721	1.93%	2.50%
INDOMOBIL PRIMA NIAGA, PT	666	1.78%	2.31%

→ will be excluded

Total unique companies = 263

Income Estimation OOT Validation

- OOT validation result shows that the model performances are quite stable (both full & nonsdk)
- Both full & nonsdk models show substantial R2, large correlation, and acceptable MAE (Mean Absolute Error)

Full	MAE(Mio)	R2	Correlation
Train	2.47	0.369	-
Test	2.37	0.361	0.613
OOT	2.58	0.346	0.622

NonSDK	MAE(Mio)	R2	Correlation
Train	2.57	0.322	-
Test	2.51	0.315	0.570
OOT	2.69	0.296	0.583

R2 Value	Interpretation
R2 < 0.02	Very weak
0.02 <= R2 < 0.13	Weak
0.13 <= R2 < 0.26	Moderate
R2 >= 0.26	Substantial

Size of Correlation (absolute)	Interpretation
r < 0.1	Very small correlation
0.1 <= r < 0.3	Small correlation
0.3 <= r < 0.5	Moderate correlation
r >= 0.5	Large correlation

**Development period : Oct'21-Dec'22, OOT period : Jan'23 - Apr-23*

**MAE is calculated by taking the average of the absolute differences between predicted & actual values*

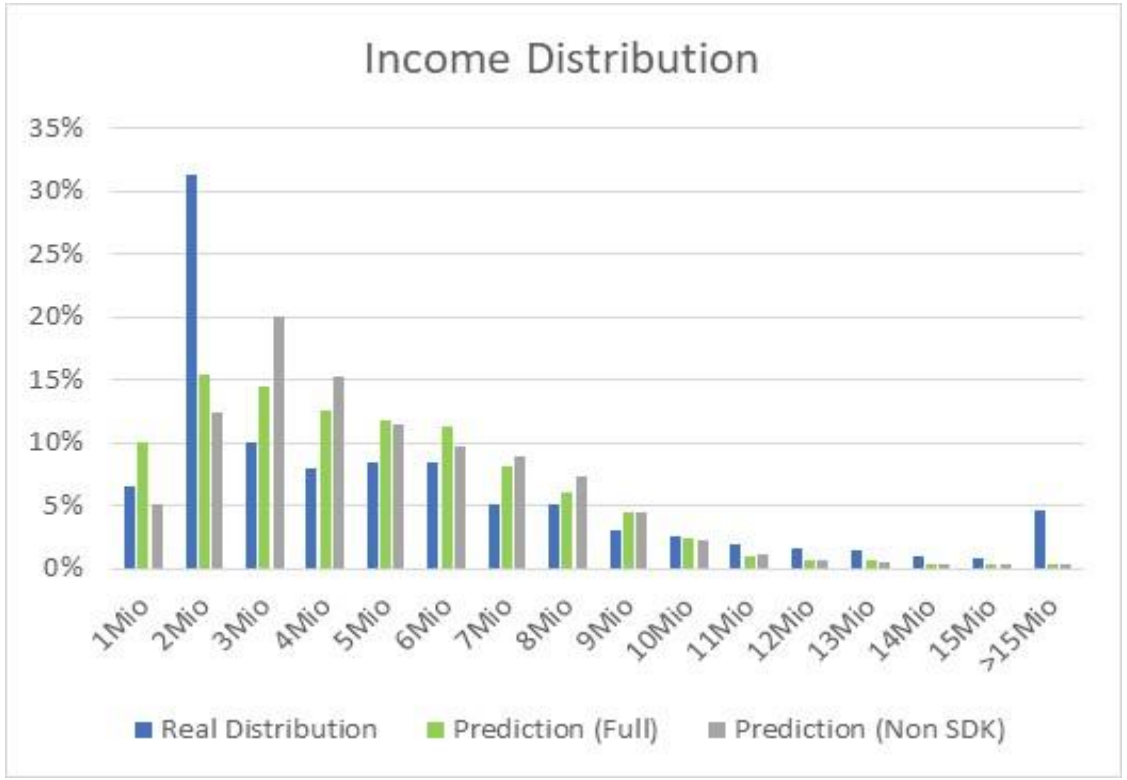
**Threshold is based on J.Cohen (1988)*

** Model method is linear regression*

Variable Waterfall & Income Distribution

bank btpn A MEMBER OF SMBC Group

Criteria	Pop	(Drop)/Add
Original variables	46	
Derived variables	69	23
Doesn't align with loan process	60	(9)
Has only 1 unique value	58	(2)
Identical rate > 95%	56	(2)
Multicollinearity	15	(41)
Significant for model	10	(5)



- Derived variables consist of variable discretization & interaction
- Full model distribution is more similar with the real income distribution compared to non sdk distribution
- High Income (>15 Mio) is difficult to predict for both full model and non sdk model

Income Model Predictor

Predictor	Used in Full Model	Used in NonSDK Model	Type of Predictor	Definition	Value Example
Age	1	1	Demographic	Age of Customer	25,26,27,…..
Brand	1	0	SDK	Brand of customer's device	Samsung, OPPO,….
Education	1	1	Demographic	Customer education level	SMP, SMA,…
Email	1	1	Demographic	Customer email	gmail, yahoo,….
Jenis_Usaha	1	1	Demographic	Customer business type	Perdagangan, Transportasi,…
LocaleDisplayLanguage	1	0	SDK	Current language selected in customer's device	English, Indonesia,….
provider_name	1	1	Demographic	Name of provider	XL, Telkomsel,…..
RamTotalSize	1	0	SDK	Ram size of customer's device	6GB, 8GB,……..
UMR	1	1	Other	minimum salary on customer's region (IDR)	3Mio, 4.9Mio
YOE_JABATAN	1	1	Demographic	Years of experience X Position	Manager- 2 Year, Manager -1 Year

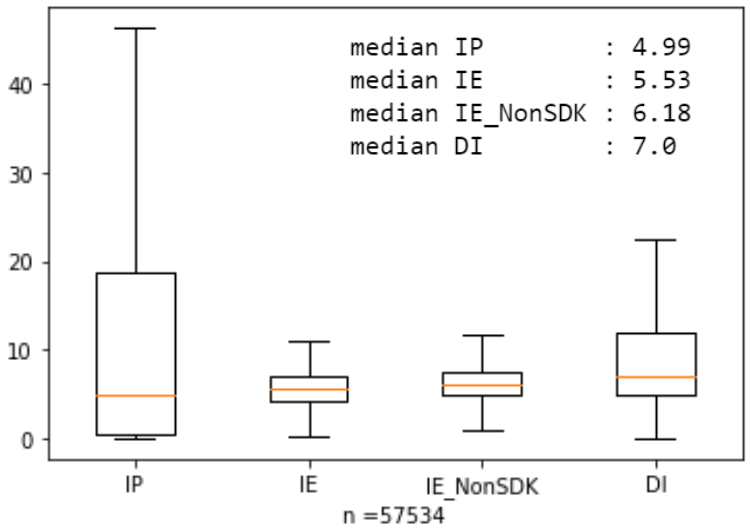
- Full Model consists of demography, external data(UMR), and SDK predictor
- Non SDK Model consist of demography and external data(UMR) predictor

Income Distribution Comparison (A,B,C,D,E,F,H)

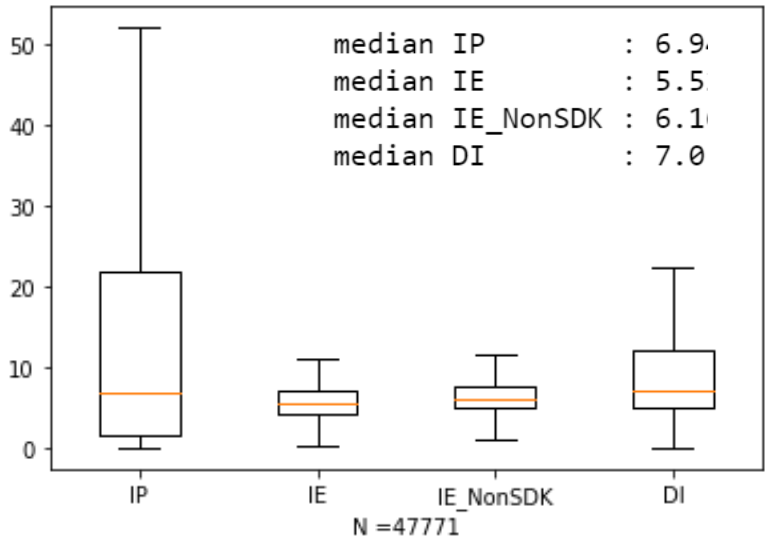
bank btpn A MEMBER OF SMBC Group

- Sample period : Nov'22 – Jan'23
- For Bureau segment the median of income estimation are higher than income proxy but lower than declared income
- For Bureau segment the median of income estimation are higher than income proxy but lower than declared income

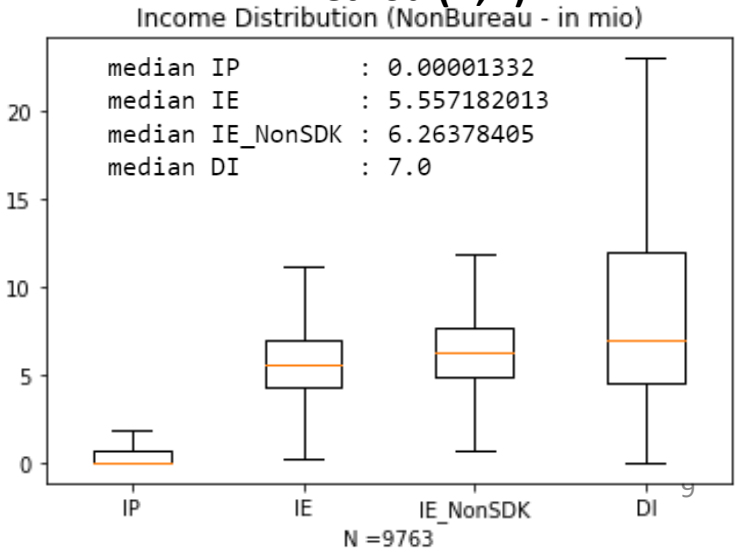
All Method (A,B,C,D,E,F,H)



Method (A,B,C,E,F,H)



Method (D,H)



**IP = Income Proxy*
IE = Income Estimation
IE_NonSDK = Income Estimation (without SDK)
DI = Declared Income

Limit Calculation Comparison (G)

Existing

Last Education	Age	Risk-based Customer Limit (in IDR)			
		LR	MR	HR	NS
S1 and Above	21 - 25	4,000,000	3,000,000	2,000,000	X
	26 – 30	3,000,000	2,000,000	1,000,000	X
	31 - 50	2,000,000	1,000,000	500,000	X
Below S1	21 - 25	2,000,000	2,000,000	1,000,000	X
	26 – 30	2,000,000	1,000,000	500,000	X
	31 - 50	1,000,000	500,000	X	X

Proposed

Risk Grade	IIR	Tenor
LR	30%	3
MR	25%	3
HR	20%	3
NS	15%	3

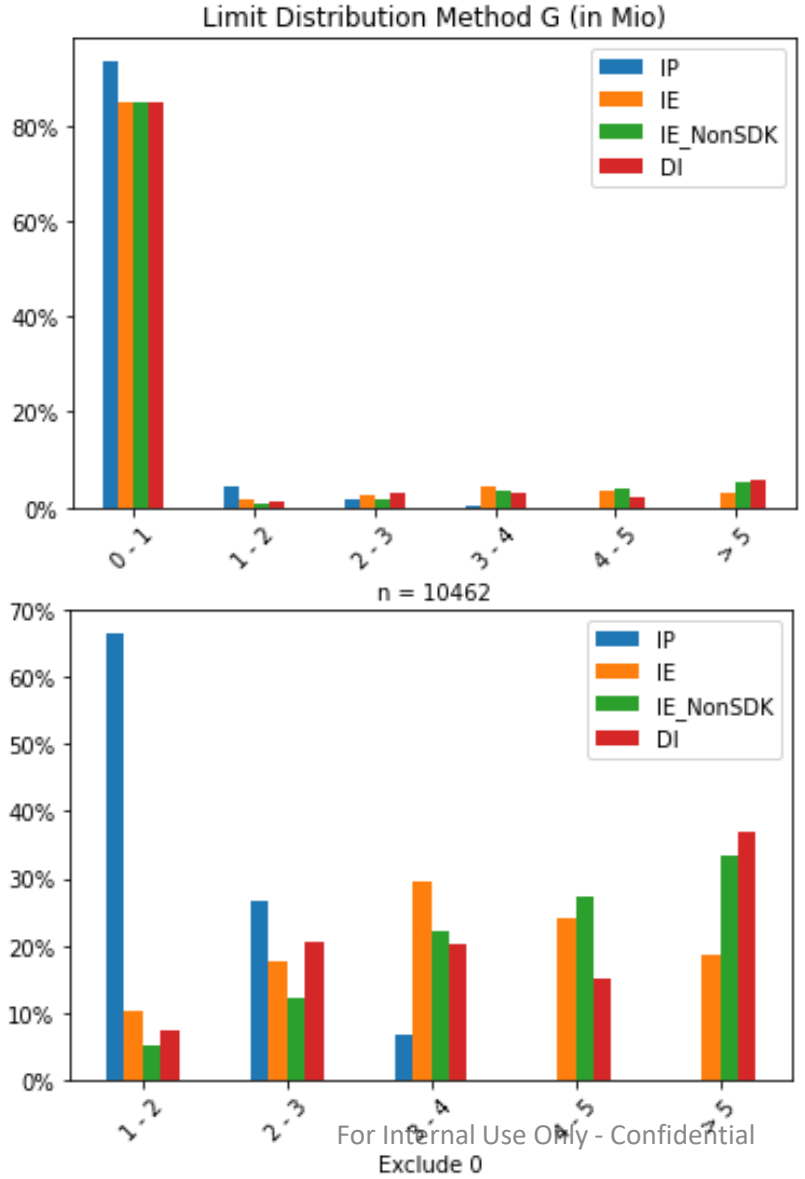
Max Credit Limit at IDR 30 Mio

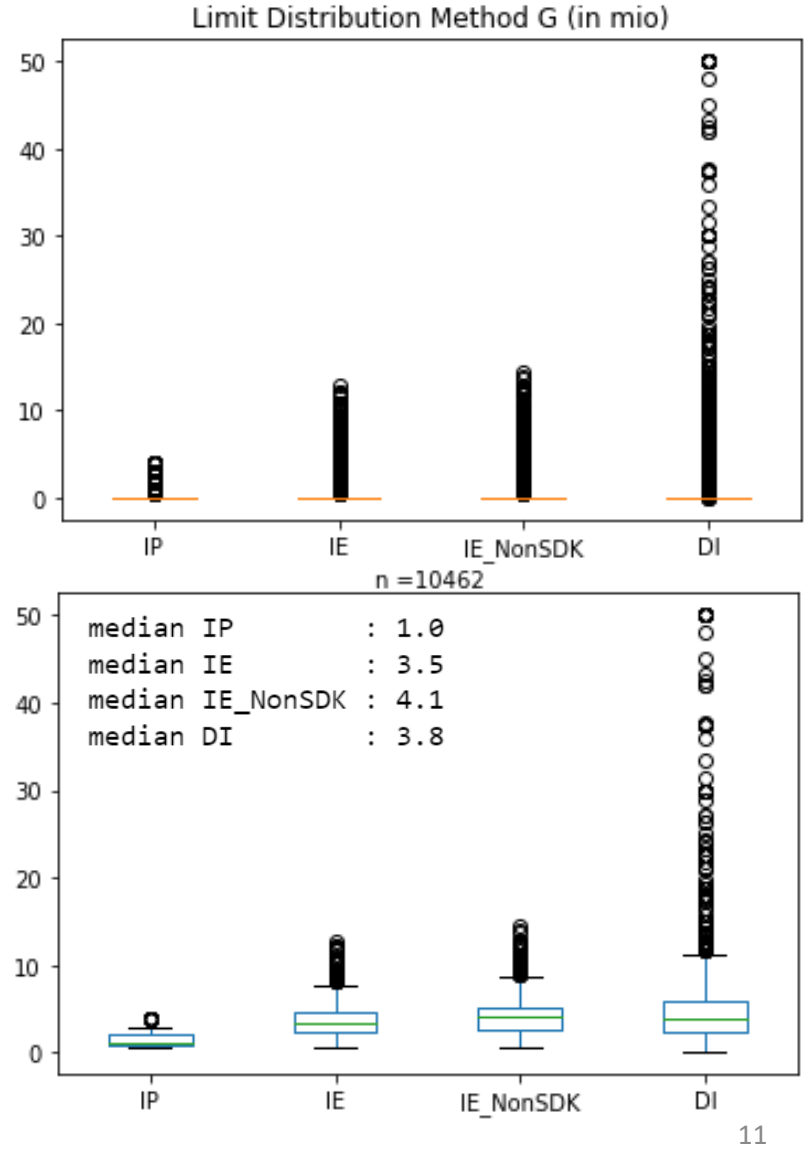
Formula : Income x IIR x Tenor

Limit Distribution Comparison

- Sample period : Nov'22 to Jan'23 based on Method G only

- Distribution of Income Estimation is quite similar to Declared Income





*IP = Income Proxy
IE = Income Estimation
IE_NonSDK = Income Estimation (without SDK)
DI = Declared Income

Swap-In & Swap-Out (In Bio)

bank btpn A MEMBER OF SMBC Group

Full Model

Decision	# Applicant	Note	IP Limit	IE Limit	Limit Increase/Decrease
APPROVE by IP	1,822		2.43	6.39	
APPROVE by IE	1,821		2.43	6.39	3.96
RE by IE	1	Swap Out	0.002	0.002	- 0.002
RE by IP	10,594		-	0.191	
APPROVE by IE	52	Swap In	-	0.188	0.19
RE by IE	10,542		-	0.003	0.00
Grand Total	**12,416**		**2.43**	**6.57**	**4.15**
					171%

	# Applicant
Recontest	**374**
Recontest due to capacity	115
Approve by IE	115
	100%

IP Decision	Decision Reason	IE Decision	# Applicant
RE	**RTP0015**	APPROVE	1
	RTP0018	APPROVE	34
	RTP0019	APPROVE	17
		RE	4
Grand Total			**56**

93%

NonSDK

Decision	# Applicant	Note	IP Limit	IE Limit	Limit Increase/Decrease
APPROVE by IP	1,822		2.43	7.34	
APPROVE by IE	1,822		2.43	7.34	4.92
RE by IE	-	Swap Out	-	-	-
RE by IP	10,594		-	0.211	
APPROVE by IE	53	Swap In	-	0.210	0.21
RE by IE	10,541		-	0.002	0.00
Grand Total	**12,416**		**2.43**	**7.55**	**5.13**
					211%

	# Applicant
Recontest	**374**
Recontest due to capacity	115
Approve by IE	115
	100%

IP Decision	Decision Reason	IE Decision	# Applicant
RE	**RTP0015**	APPROVE	1
	RTP0018	APPROVE	35
	RTP0019	APPROVE	17
		RE	3
Grand Total			**56**

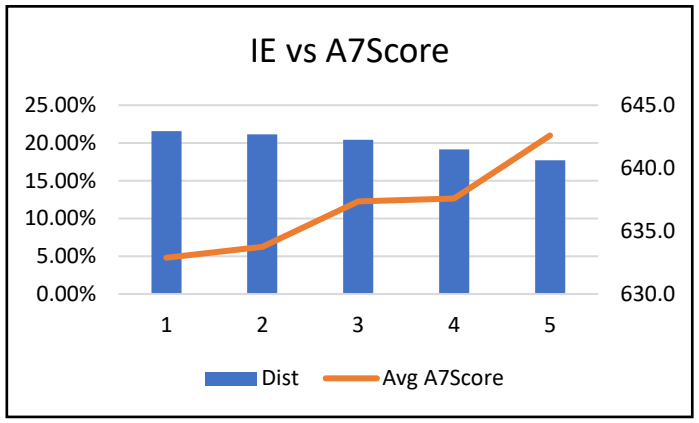
95%

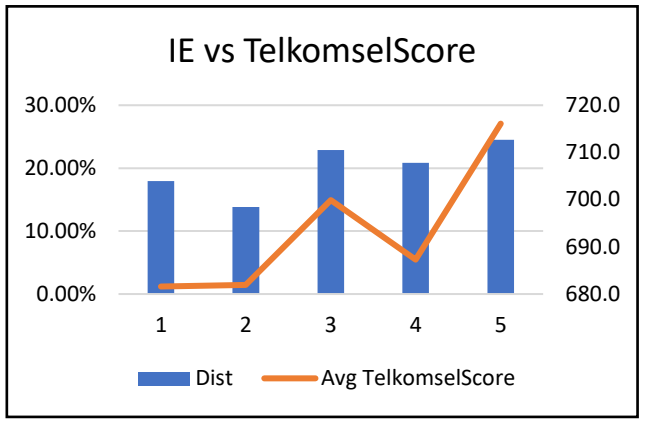
- Sample period : Nov'22 to Jan'23 based on Method G
- Income Estimation (both Full & NonSDK) show slight approval rate increase & significant credit limit
- From 56 Applicants which rejected by IP due to capacity below threshold, **93-95%** of it approved by IE

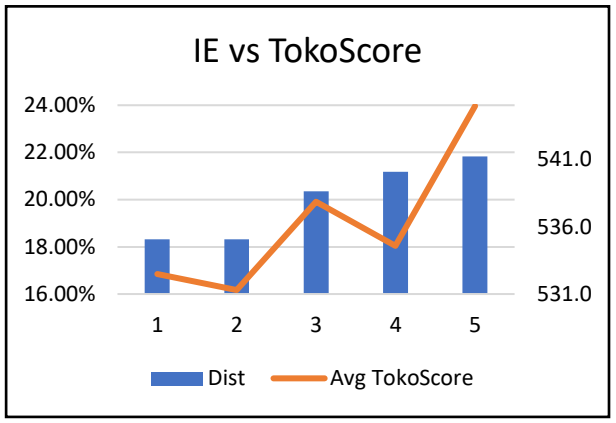
*RTP0015, RTP0018, RTP0019 are reject reason due to capacity below threshold

Income Estimation vs Credit Score

- Sample period : Nov'22 to Jan'23 based on Method G
- The Graph show the correlation between Income quintile and average score
- The higher the Income Estimation, the higher the average score







N= 10.4K

N= 1.6K

N= 1K