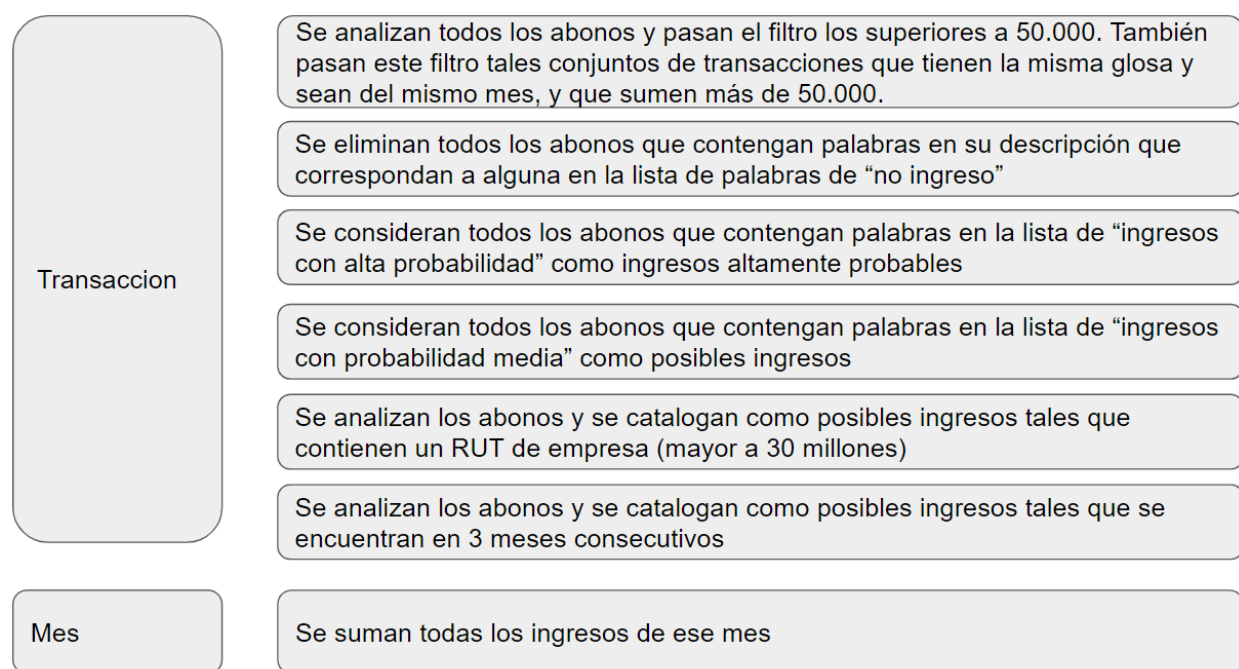


Algoritmo Income 2.0

Cristián Fazio

Resumen:



Definición de transacción de ingreso

Una transacción debe cumplir con las siguientes condiciones para ser considerada una transacción de ingreso u abono (de aquí en adelante). Antes que nada, se van a sumar los montos de las transacciones que tengan la misma glosa en el mismo mes, de esta forma transacciones que se repiten el mismo mes pasan por los filtros siguientes en conjunto. En primer lugar el monto tranzado debe ser superior a 50.000 (parámetro ajustable), dado que los montos menores a este probablemente correspondan a ingresos informales o transacciones menores que no corresponden a ingresos reales.

Después de filtrar los abonos con el filtro anterior, se procede a un análisis de lenguaje basado en la identificación de palabras y de patrones en las descripciones de cada transacción de abono, el cual consiste de tres etapas:

Primera Etapa

Se utiliza una lista de palabras (`not_income_words`, parámetro ajustable) que consiste de palabras que al presentarse en la descripción de un abono indican que el abono en cuestión no es un ingreso. Una lista tentativa simplificada (pero de ninguna forma exhaustiva) es la siguiente:

```
not_income_words = ["ahorro", "inversion", "fintual", "rescate", "modelo", "afp",  
                    "porciento", "ife", "devolucion"]
```

Por ejemplo:

```
"IFE UNIVERSAL"  
"PAGO DIEZ POR CIENTO AFP MODELO"  
"PAGO TRES DIEZ PORCIENTO HABITAT"  
"EVOLUCION:UBER EATS HELP.UB"
```

Al identificar una de estas palabras en la descripción de un abono, la transacción en cuestión es eliminada del análisis de las siguientes etapas. Es importante notar que este método permite la identificación de otros tipos de abonos, ya que es posible identificar abonos que contengan palabras relacionadas a inversiones o relacionadas a beneficios sociales.

Segunda Etapa

En esta etapa se identifica la probabilidad de los abonos restantes de ser un ingreso real mediante un análisis similar al de la primera etapa. En primer lugar se definen dos

listas de palabras, la primera (`high_income_words`, parámetro ajustable) para palabras que se encuentran comúnmente en descripciones de abonos que indican un alta probabilidad de que el abono en cuestión sea un ingreso real. Una posible lista podría ser la siguiente:

```
high_income_words = ["ingreso", "remuneracion", "sueldo"]
```

Por ejemplo:

```
"PAGO:DE SUELDOS 0765252903"
```

Similarmente, la segunda lista (`medium_income_words`, parámetro ajustable) contiene palabras que indican que el abono en cuestión tiene una probabilidad media de ser un ingreso real. Un ejemplo de una lista así podría ser a siguiente:

```
medium_income_list = ["spa", "deposito", "efectivo", "sociedad",  
                      "fundacion", "ingenieria", "transportes"]
```

Por ejemplo:

```
"Transf. de TRANSPORTES TONINI"
```

```
"TEF DE DE RUT 76754263-1DEPOSITO EN EFECTIVO CAJA VECINA"
```

Al identificar que alguna de estas palabras de alguna lista se encuentra en la descripción de un abono, la transacción se cataloga como un ingreso de probabilidad alta o media (dependiendo del caso) y no se considera para la siguiente etapa. Adicionalmente, se consideran abonos de ingreso con probabilidad media tales abonos que en su descripción contengan un RUT mayor a 30 millones (parámetro ajustable).

Por ejemplo:

```
"TEF DE DE RUT 76754263-1"
```

Tercera Etapa

Esta etapa hace un análisis en base a la recurrencia de transacciones. Simplemente, si es que la misma descripción de un abono se repite por lo menos en 3 (parámetro ajustable) meses consecutivos, se considera que todas las transacciones con esa descripción son un ingreso de probabilidad media. Esto es necesario ya que es común ver depósitos que en su descripción tienen el nombre de una empresa y **no** tienen el RUT o alguna palabra que los identifique como empresa, por lo que este método cubre a todas las personas que trabajan recurrentemente en una empresa.

Anexos

Anexo A: Parámetros ajustables del algoritmo

Filtro Inicial:

- monto_minimo_tranzado

Primera Etapa:

- not_income_words

Segunda Etapa:

- high_income_words
- medium_income_words
- rut_minimo

Tercera Etapa:

- consecutividad

Anexo B: Diagrama de Flujo de calificación de un abono

