

G13: Licitaciones públicas - ChileCompra

Andrés Carvallo

Camilo Rojas

Jonathan Wai Lam Lee

Feedback Hito 2

- Puede mostrar ejemplos reales
- La relación entre los experimentos y el objetivo no está clara
- No está claro el problema y el objetivo

Agenda

1. Introducción
2. 3 experimentos
3. Conclusion
4. Trabajo futuro

Ciclo de vida de una licitación



UrlLicitacion:

<http://www.mercadopublico.cl/Procurement/Modules/RFB/StepsProcessAward/PreviewAwardAct.aspx?qs=Ldp8tkLKr00+pAjm3U5za6iB/oOAjIVYVc67zQvQhQA=>

Oportunidad para DM

- Nov 2016 Region Metropolitana
 - 3640 licitaciones
 - 13582 productos licitados
 - 222 organizaciones públicas
 - > 1761 proveedores (69,2% micro, pequeña y mediana [1])
-
- $3640 / 30 = 121$ licitaciones por dia
 - 452 productos licitados por dia

Problema general

- No es factible revisar >452 productos/día
- Atraer más oferentes
- Evaluar la calidad de licitación

Ejemplo de aplicación

$$\text{rate}_{\text{proveedor}}(\text{licitación}) = f_{\text{proveedor}}(\text{licitación.atr1}, \text{licitación.atr2}, \dots, \text{licitación.atrn})$$

- Sistema recomendador

Que es una buena licitación?

- Atributos? (adjudicación, precio adjudicado, número de oferentes)
- Perspectivas? (general, proveedor, organización pública)
- Experimentar con 2 aspectos
 - a. Adjudicacion
 - b. Proveedor

E1: Trazas de licitaciones

- ~20% desierta (3000 licitaciones, 7000 productos)
- Qué pasa con las licitaciones desiertas?
- Trazas de licitaciones

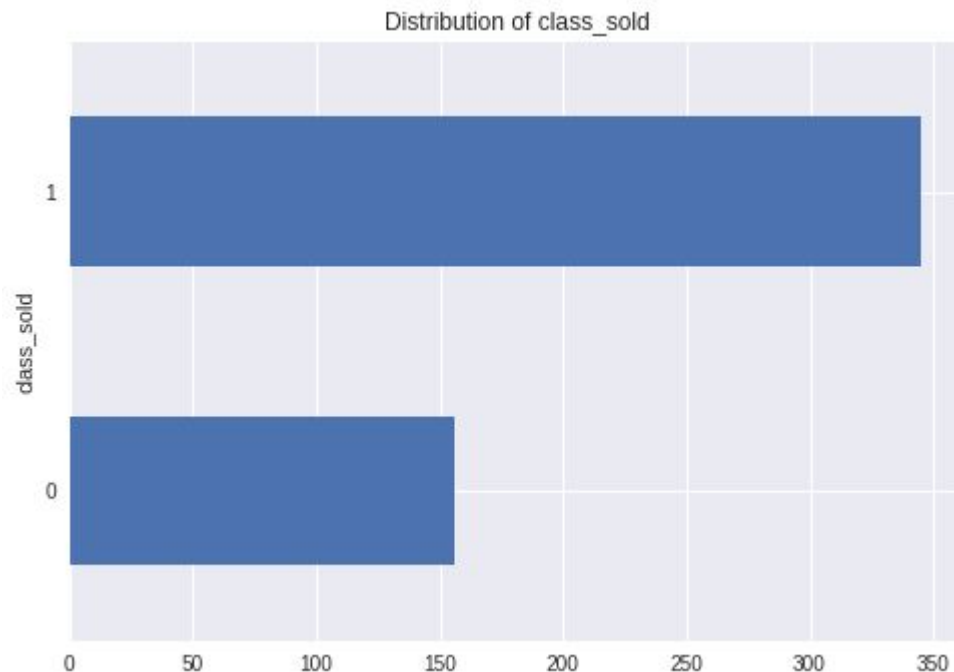
| TID | Estado | Publicacion | Cierre | Producto | Organización |
|-------------------------------|--------|-------------|------------|----------------------------|--------------------------------|
| 2592-622-1E16 | D | 2016-11-03 | 2016-11-14 | Oficina para 14 fotógrafos | Ilustre Municipalidad de Flaco |
| 2592-704-1E06 | D | 2016-11-29 | 2016-12-03 | Oficina para 14 fotógrafos | Ilustre Municipalidad de Flaco |
| | | | | | |

E1: Preguntas de investigación

- Transformar desierta (mala) a adjudicada (buena)
- Cuales son las características de trazas que terminan con estado adjudicada?

E1: Experimentos

- Desequilibrio: $350/500 = \sim 70\%$
- Train (Original, Oversample)
- Test (Original, NoDuplicado)
- 4 variaciones



E1: Resultados

- Train (Original), Test (NoDuplicado)
- Mejor que Dummy (Stratified)
- AdaBoost es DT configurado para ruido

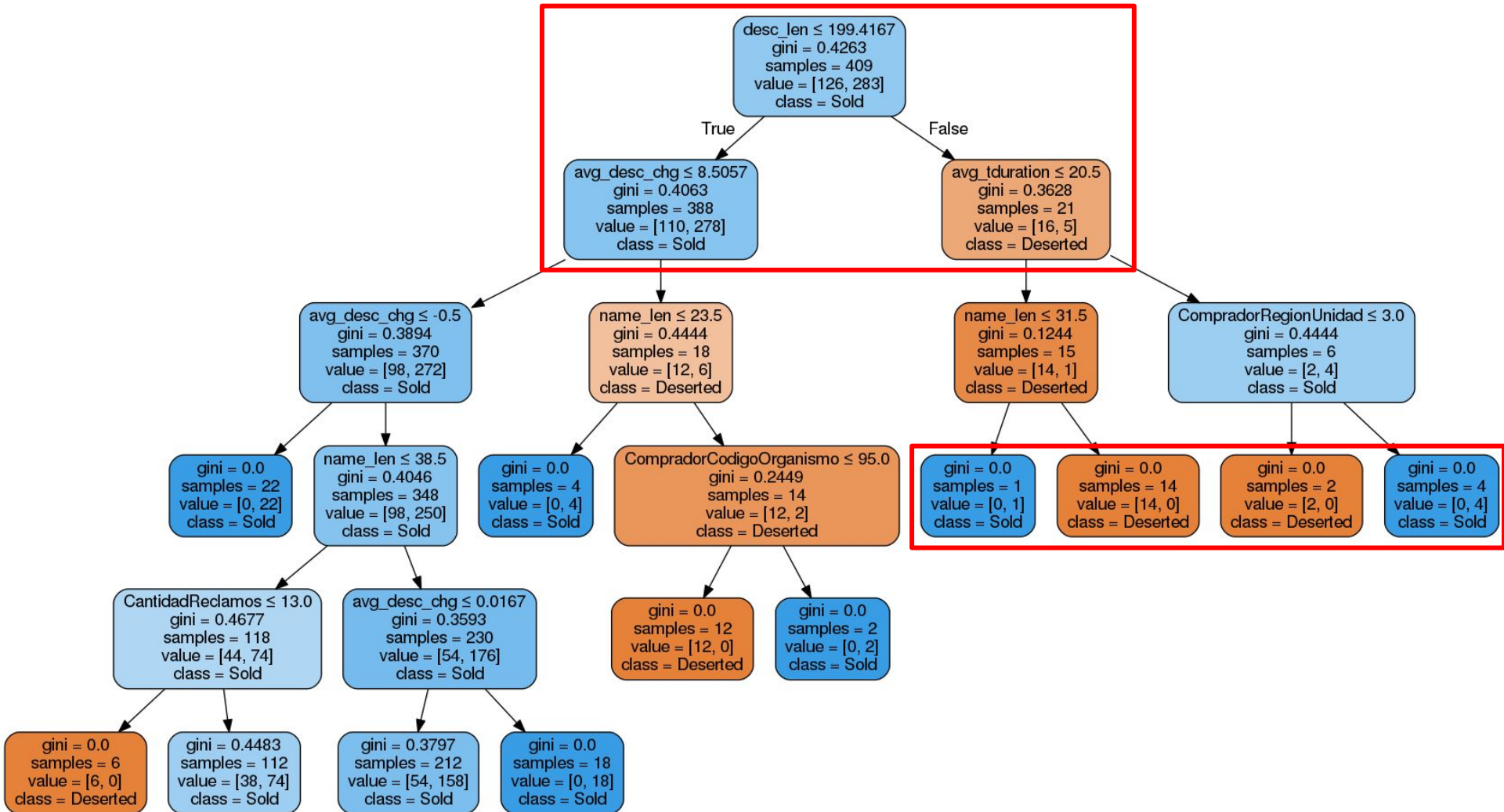
| Dummy (Stratified) | | precision | recall | f1-score | support |
|--------------------|---|-----------|--------|----------|---------|
| | 0 | 0.22 | 0.28 | 0.25 | 25 |
| | 1 | 0.77 | 0.70 | 0.73 | 84 |
| avg / total | | 0.64 | 0.61 | 0.62 | 109 |

| Nearest Neighbors | | precision | recall | f1-score | support |
|-------------------|---|-----------|--------|----------|---------|
| | 0 | 0.41 | 0.48 | 0.44 | 25 |
| | 1 | 0.84 | 0.80 | 0.82 | 84 |
| avg / total | | 0.74 | 0.72 | 0.73 | 109 |

| Decision Tree | | precision | recall | f1-score | support |
|---------------|---|-----------|--------|----------|---------|
| | 0 | 0.70 | 0.28 | 0.40 | 25 |
| | 1 | 0.82 | 0.96 | 0.89 | 84 |
| avg / total | | 0.79 | 0.81 | 0.77 | 109 |

| Random Forest | | precision | recall | f1-score | support |
|---------------|---|-----------|--------|----------|---------|
| | 0 | 0.71 | 0.20 | 0.31 | 25 |
| | 1 | 0.80 | 0.98 | 0.88 | 84 |
| avg / total | | 0.78 | 0.80 | 0.75 | 109 |

| AdaBoost | | precision | recall | f1-score | support |
|-------------|---|-----------|--------|----------|---------|
| | 0 | 0.50 | 0.48 | 0.49 | 25 |
| | 1 | 0.85 | 0.86 | 0.85 | 84 |
| avg / total | | 0.77 | 0.77 | 0.77 | 109 |



E1: Reflexiones

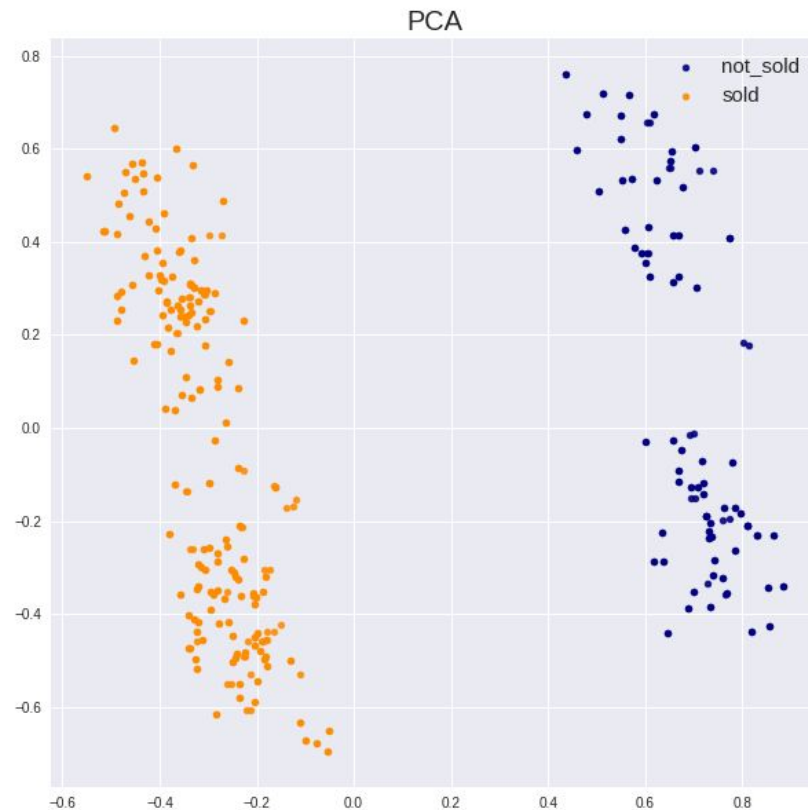
- Es buena práctica re-publicar una licitación
- Se puede aplicar clasificación
- Quizas debería usar clustering en vez de clasificación

E2: Preguntas de investigación

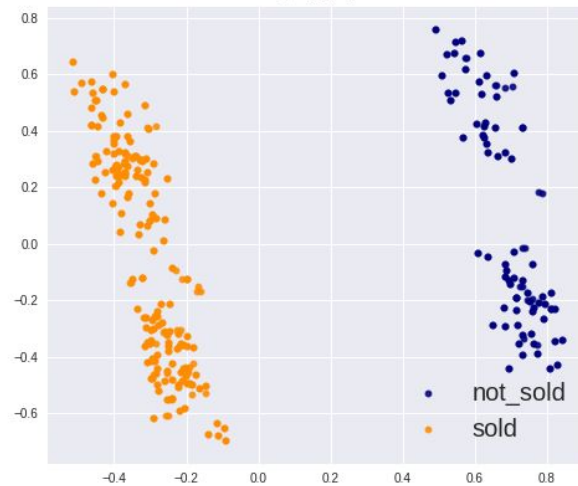
- Podemos aplicar clustering para verificar que la clase *class_sold* tiene sentido?

E2: Clustering

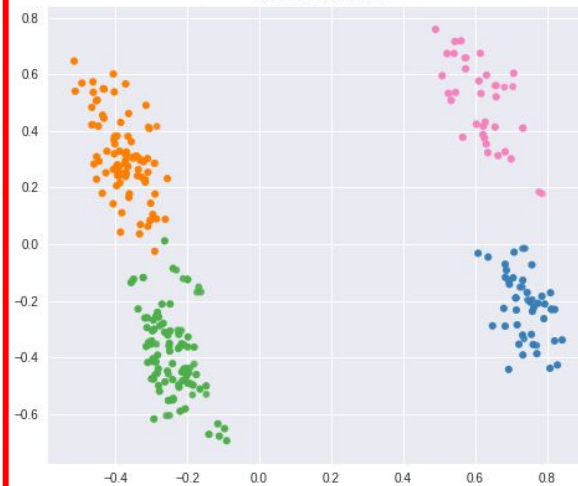
- Verificar con clustering
- Visualizar trace_df (10 dim) PCA
- Usar *clase_sold* como label



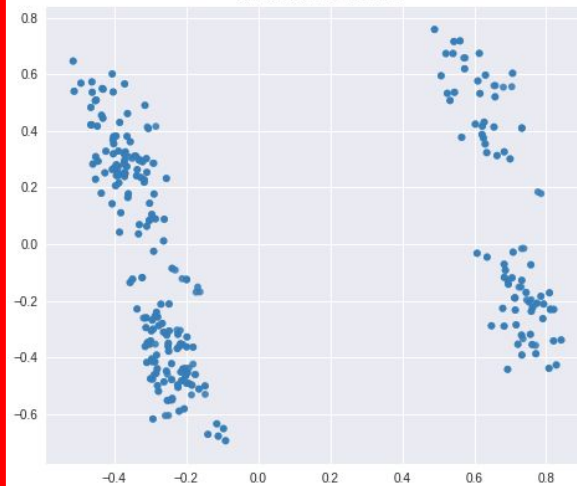
PCA



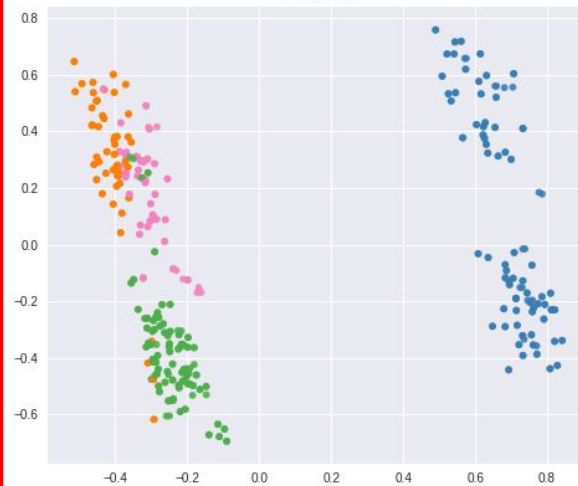
KMeans



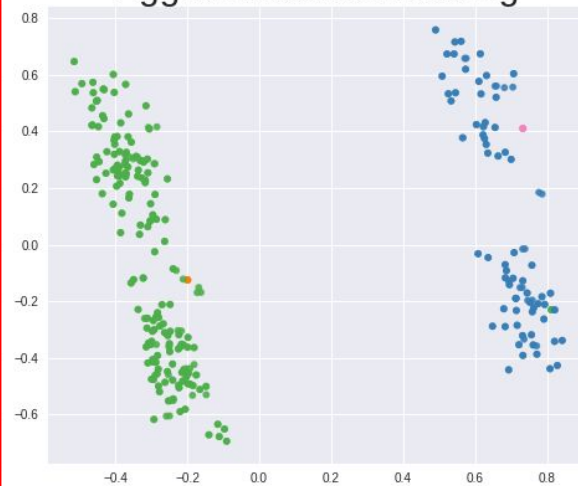
MeanShift



Ward



AgglomerativeClustering



DBSCAN



E2: Reflexiones

- Calidad de los clusters
- Verificación de la intuición anterior
- Significado de los clusters

E3: Experimento clasificación proveedores

- Para este experimento queremos determinar qué features son relevantes para que un proveedor aumente sus probabilidades tener éxito en una licitación.
- Primero que nada un proveedor puede ofrecer muchos productos y una licitación también puede ser abastecida de uno o más productos.

$$\text{proveedor}_i = \{ \text{producto}_1, \text{producto}_2, \dots \}$$



E3: Probabilidad de adjudicación de licitación

- Como un proveedor puede participar de una licitación con varios productos.
- Consideramos que una empresa que tiene una mayor cantidad de productos ofrecidos respecto al total de productos ponderados del dataset , es una empresa con más chances de adjudicarse licitación (exitosa)

$$proveedor_i = \{ producto_1, producto_2, \dots \}$$

$$\frac{licitacion(1)}{licitacion(1)+licitacion(2)} * \frac{licitacion^{Pi}(1)}{licitacion(1)} + \frac{licitacion(2)}{licitacion(1)+licitacion(2)} * \frac{licitacion^{Pi}(2)}{licitacion(2)}$$

E3: Experimento clasificación proveedores

- Hacemos el siguiente supuesto:
- **Empresa exitosa:** un alto porcentaje de productos respecto del total de productos del dataset
- **Empresa medianamente exitosa:** porcentaje intermedio
- **Empresa poco exitosa:** porcentaje bajo

```
# 0 no exitosa, 1 medianamente exitosa, 2 exitosa
def tramo_success(row):
    if 0 < row['prob'] < 0.003:
        return 0
    if 0.003 <= row['prob'] < 0.005:
        return 1
    else:
        return 2
```



Por qué esta división?

Porque los datos tienen una distribución de esta forma...

E3: Experimento clasificación proveedores

- Ocupamos clasificación con Bayes, Decision Tree, Dummy, K-Neighbors
- Consideramos como variable target el tramo al que pertenecería el proveedor de acuerdo a su probabilidad de éxito (2 exitosa, 1 mediana, 0 no exitosa)
- Features considerados:
 - **Región (1 metropolitana, 0 otras regiones)**
 - **Monto Licitación**
 - **Cantidad (Volumen)**
 - **Precio unitario**



| Dummy Classifier | | | | |
|------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.18 | 0.19 | 0.18 | 169 |
| 1 | 0.14 | 0.12 | 0.13 | 144 |
| 2 | 0.73 | 0.73 | 0.73 | 845 |
| avg / total | 0.58 | 0.58 | 0.58 | 1158 |

| Decision Tree Classifier | | | | |
|--------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.32 | 0.33 | 0.32 | 169 |
| 1 | 0.24 | 0.25 | 0.24 | 144 |
| 2 | 0.79 | 0.78 | 0.79 | 845 |
| avg / total | 0.65 | 0.65 | 0.65 | 1158 |

| Gaussian Naive Bayes Classifier | | | | |
|---------------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.50 | 0.07 | 0.12 | 169 |
| 1 | 0.40 | 0.01 | 0.03 | 144 |
| 2 | 0.74 | 0.99 | 0.84 | 845 |
| avg / total | 0.66 | 0.73 | 0.64 | 1158 |

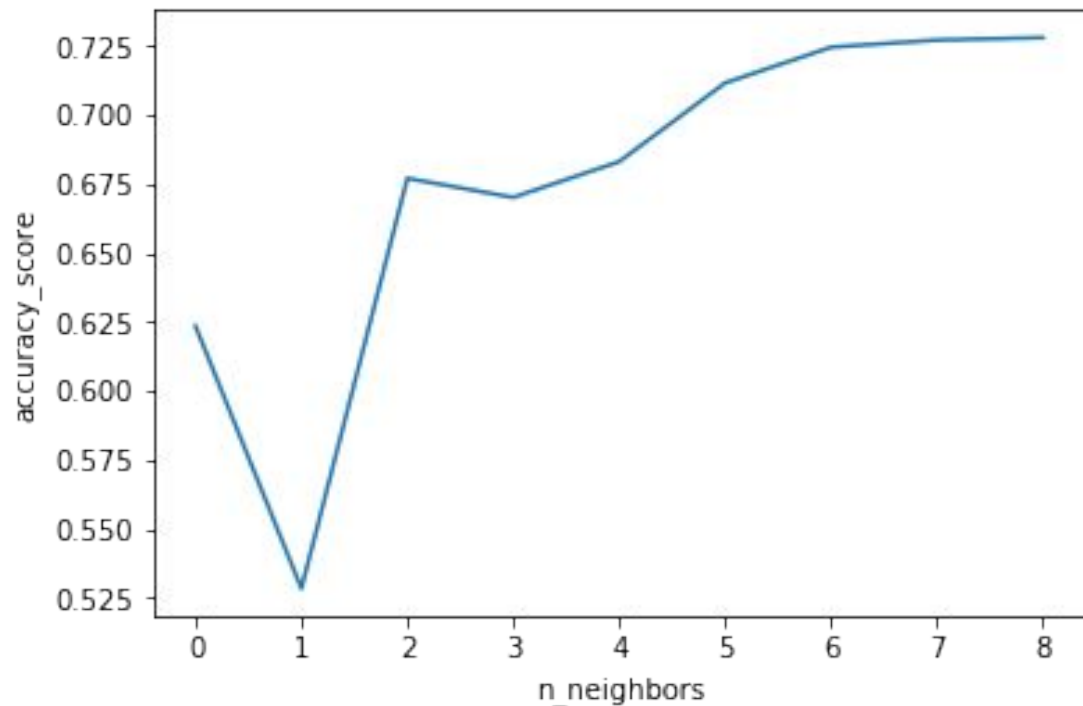
| KNeighbors Classiffier | | | | |
|------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.36 | 0.25 | 0.29 | 169 |
| 1 | 0.36 | 0.12 | 0.19 | 144 |
| 2 | 0.77 | 0.90 | 0.83 | 845 |
| avg / total | 0.66 | 0.71 | 0.67 | 1158 |

E3: Resultados

Mejor clasificador es:

K-NEIGHBORS CLASSIFIER

E3: Resultados



Vemos que al aumentar la cantidad de vecinos aumenta el **accuracy score**.

E3: Resultados y observaciones del clasificador

| | No exitosa (0) | Mediana exitosa (1) | Exitosa (2) |
|------------------|----------------|---------------------|-------------|
| No exitosa (0) | 32 | 9 | 128 |
| Medianamente (1) | 12 | 17 | 115 |
| Exitosa (2) | 33 | 18 | 794 |

El algoritmo de clasificación tiene un sesgo por clasificar empresas exitosas por sobre no exitosa y medianamente exitosas.

En definitiva una empresa que **NO ES EXITOSA** para **aumentar las chances de adjudicarse una licitación** debe tener en consideración **ofertas de la región metropolitana, el precio de la licitación y el volumen de producción exigido**, mismo que a veces está fuera de su alcance...

Conclusion

- Fenómeno de trazas de licitaciones
- Sugerencia de re-publicación
- Definir el éxito de un proveedor

Trabajo futuro

- Seguir con la investigación de trazas de licitaciones
- Otras perspectivas:
 - Organización: efecto de sus características, e.g., número de reclamos
- Otras técnicas:
 - NLP para descripción
 - Word2Vec para establecer relaciones entre productos para resolver problemas de granularidad

END

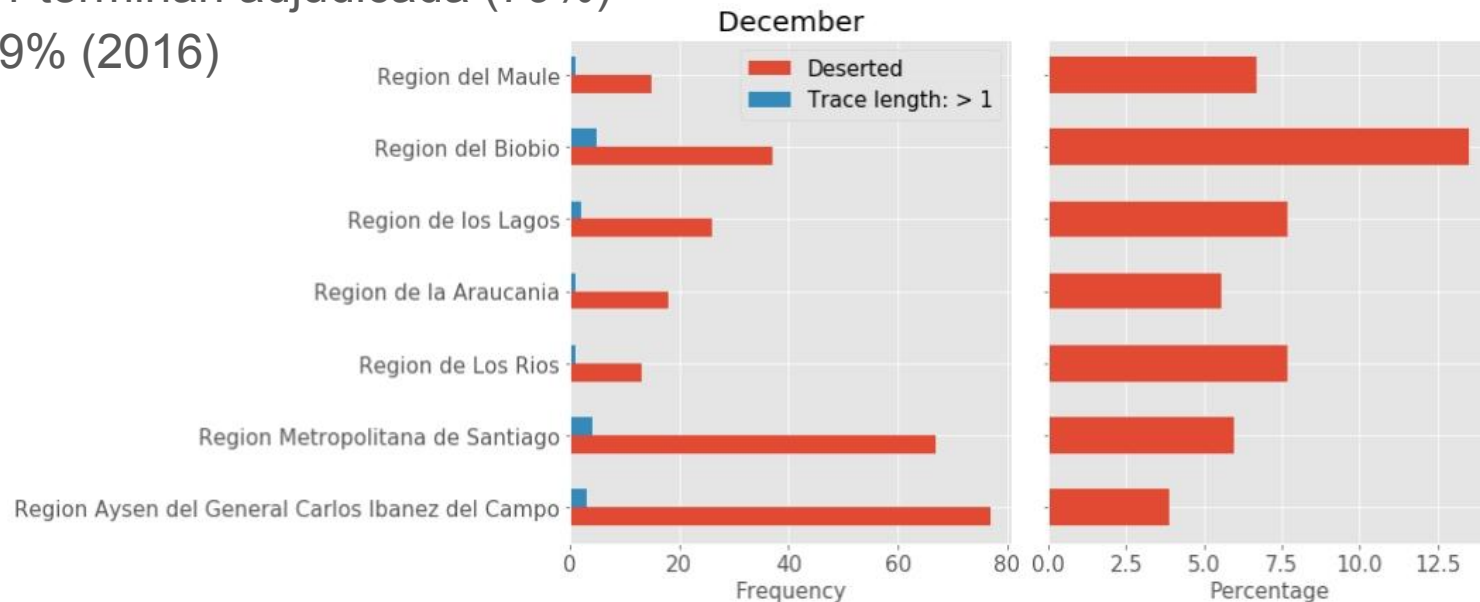
Desafíos en la calidad de datos

- Agrupación de productos licitados
- Falta granularidad en la categorización de productos
- Nivel de detalle no es uniforme

| CodigoProducto | Descripcion | AMontoUnitario | ACantidad | Proveedor |
|----------------|--|----------------|-----------|-----------------------------|
| 6012506 | Juguetes de navidad... | 11684030 CLP | 1.0 | Importaciones Maya Limitada |
| 6012506 | Implementos deportivos para taller de relajacion | 1230900 CLP | 1.0 | Manuel Emilio Munoz... |

E1: Estadísticas

- Todo el 2016
- Dec: 17 / 319 trazas tiene largo > 1 (5%)
- 11 terminan adjudicada (79%)
- 69% (2016)



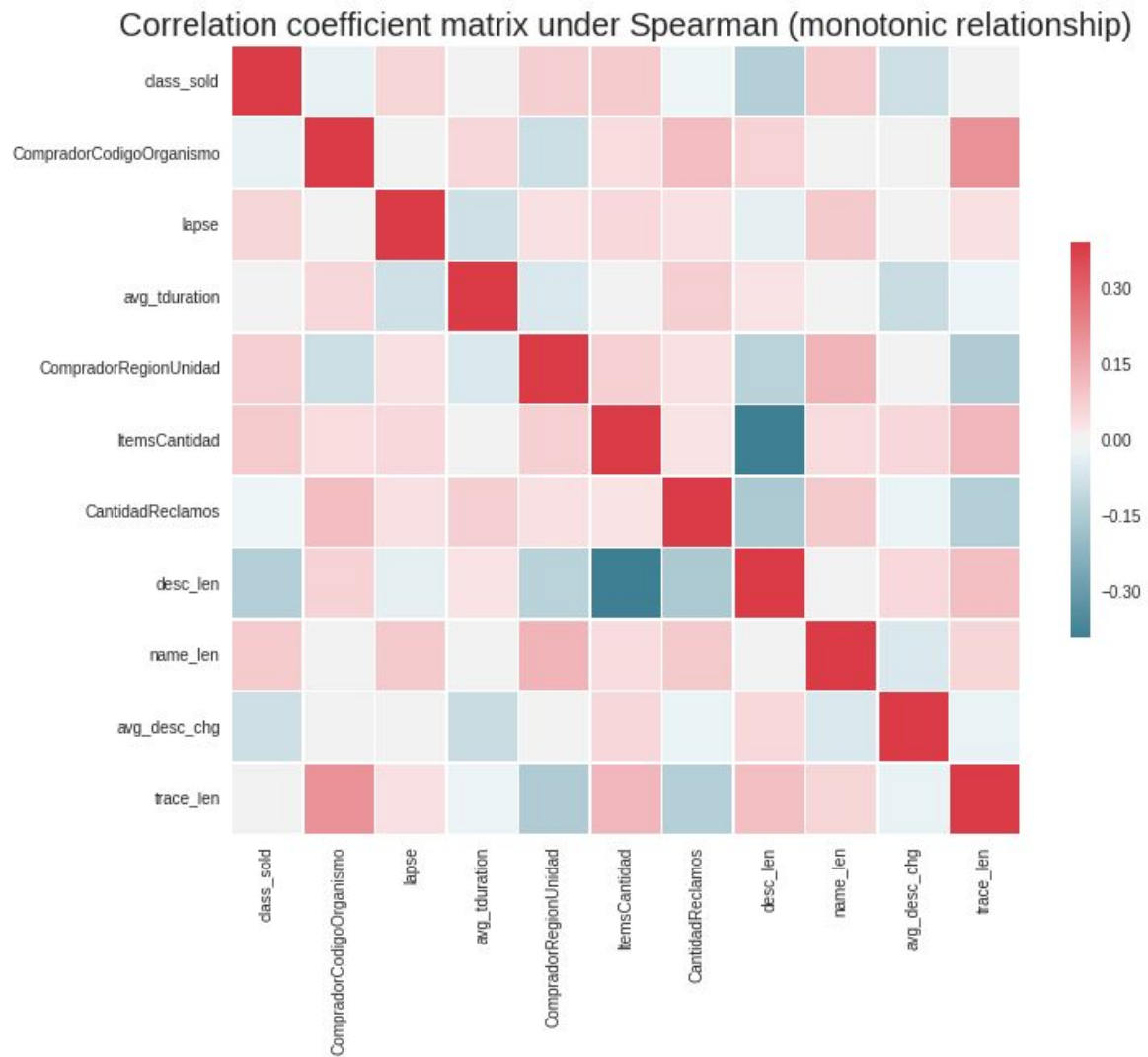
E1: Pre-procesamiento

- Eliminar ruido e información con varianza baja
- Eliminar licitaciones con un producto
- No considerar licitaciones relacionadas a productos médicos
- Para cada licitación desierta, buscamos otras licitaciones con:
 - Nombre
 - La misma organización
 - Los mismos productos licitados (tipo y cantidad)
 - Fechas
- Ampliar al dataset de 2016

Dimensiones de trace_df

- 10 dim
- Class_sold, CompradorCodigoOrganismo, lapse (entre licitaciones), avg_tduration, CompradorRegionUnidad, ItemsCantidad, CantidadReclamos, desc_len, avg_desc_chg, trace_len

E1: Correlacion



E2: Clustering

- Método del codo para encontrar optimal k
- Alta dimension problema

