Javier Ortiz Montenegro

**Tarea #1**

* **Detalles metodológicos:**

En la realización de esta tarea se han utilizado nueve conjuntos de datos, todos ellos datos de un mes (Junio de 2011) de transacciones de acciones negociadas en el NASDAQ. Tres de ellos son acciones de gran capitalización bursátil. Tres de capitalización mediana. Y por último tres de capitalización pequeña.

El procedimiento ha sido el mismo para cada uno de los conjuntos de datos. Carga del dataset en R, ejecución de un algoritmo consistente en un “for” con cadenas de “if” para de esta manera clasificar las transacciones en compras o ventas. Tanto para el “tick rule” como para el “quote rule” y “Lee y Ready”.

Una vez almacenada la clasificación en una columna por cada uno de los métodos, se ha calculado la precisión diaria y agregada por algoritmo y activo, así como también el volumen de compra y venta y su precisión.  
Una vez calculados todos estos valores se procede a mostrar las tablas con los resultados pertinentes.

* **Resultados:**

Los NaN son los fines de semana, en los cuales el mercado bursátil se encuentra cerrado.

**Tabla I:**Esta tabla recoge el porcentaje de transacciones correctamente clasificadas por cada algoritmo y cada día

| tick\_rule| quote\_rule| Lee y Ready|

|---------:|----------:|-----------:|

| 0.7866813| 0.8262889| 0.8708841|

| 0.7803296| 0.7830377| 0.8403215|

| 0.8043982| 0.6931776| 0.7985606|

| NaN| NaN| NaN|

| NaN| NaN| NaN|

| 0.8094080| 0.6988070| 0.7900976|

| 0.8012366| 0.7417400| 0.8160245|

| 0.7965879| 0.7518513| 0.8249915|

| 0.7931698| 0.7923146| 0.8566105|

| 0.7899760| 0.7162053| 0.8008129|

| NaN| NaN| NaN|

| NaN| NaN| NaN|

| 0.7787266| 0.6533523| 0.7493426|

| 0.8118627| 0.8002497| 0.8591585|

| 0.7776577| 0.7360947| 0.8127823|

| 0.7697505| 0.7288957| 0.7959420|

| 0.8079046| 0.7384025| 0.8012492|

| NaN| NaN| NaN|

| NaN| NaN| NaN|

| 0.7871927| 0.6531671| 0.7309249|

| 0.7683364| 0.7846410| 0.8399586|

| 0.7977209| 0.7371918| 0.8152500|

| 0.7838352| 0.6980649| 0.7749021|

| 0.8016989| 0.7052884| 0.7830998|

| NaN| NaN| NaN|

| NaN| NaN| NaN|

| 0.7895803| 0.7117855| 0.7828028|

| 0.7816938| 0.7371160| 0.8076163|

| 0.7587502| 0.7075465| 0.7916457|

| 0.7813809| 0.7333407| 0.7965880|

En esta primera tabla se puede apreciar como la precisión varía según el día analizado, pero suele situarse entre un 70% y 85%. También muestra que el método de clasificación más preciso va cambiando, pero el que clasifica con mayor precisión la mayoría de los veces parece ser Lee y Ready y método con menor precisión quote rule.

**Tabla II:**

En la segunda tabla se recogen las medias y desviaciones típicas diarias por activo.

| | Media| Desv. Típica|

|:---------|---------:|------------:|

|stock\_1 | 0.7947515| 0.0356070|

|stock\_2 | 0.8074872| 0.0286783|

|stock\_3 | 0.7385408| 0.0565866|

|stock\_101 | 0.8283042| 0.0396439|

|stock\_102 | 0.7216198| 0.0633223|

|stock\_103 | 0.7182541| 0.0645454|

|stock\_201 | 0.7690061| 0.0665138|

|stock\_202 | 0.7902772| 0.0589594|

|stock\_203 | 0.8171234| 0.0611356|

Esta segunda tabla muestra como aún cuando las medias se mantienen entre los mismos rangos de manera independiente al volumen de transacciones de los activos analizados, no es así para la desviación típica. Por lo que puede parecer en la tabla la desviación típica se reduce para los activos con mayor cantidad de transacciones.

**Tabla III:**

Esta última tabla recoge una medida agregada de precisión (Volumen correctamente clasificado sobre el volumen total de transacciones) por activo y por algoritmo de clasificación.

| | tick\_rule| quote\_rule| Lee y Ready|

|:---------|---------:|----------:|-----------:|

|stock\_1 | 0.9890726| 0.7436297| 0.9844394|

|stock\_2 | 0.9909858| 0.9480532| 0.9943004|

|stock\_3 | 0.9862379| 0.8342911| 0.9877097|

|stock\_101 | 0.9828568| 0.8280181| 0.9820450|

|stock\_102 | 0.9498308| 0.9179953| 0.9602229|

|stock\_103 | 0.9661177| 0.8640774| 0.9529480|

|stock\_201 | 0.9451220| 0.9370432| 0.9741801|

|stock\_202 | 0.9378967| 0.9042450| 0.9600200|

|stock\_203 | 0.9712744| 0.8737172| 0.9735858|

Se puede apreciar en la tabla como los algoritmos con una precisión excepcionalmente buena son Tick rule y Lee y Ready, siendo este último ligeramente más preciso.

* **Conclusiones:**

Si comparamos la tabla I y la tabla III vemos que la precisión varía bastante según como la observemos (día a día o con una medida agregada por activo). Esto podría ser debido a que la tabla III tiene en cuenta el volumen a la hora de realizar el cálculo, es decir, cuando una transacción tiene un gran volumen la precisión de los métodos de clasificación aumenta.

Esta hipótesis se refuerza si observamos el comportamiento de la desviación típica para activos con mayor volumen frente a los de menor volumen en la tabla II.