הגדרת פונקציות

1. פונקציה ראשונה - StockDataBuilder

פונקציה זו תקל קבצי csv בהם קיימות החברות אשר אנו נכלול במחקרינו.

* בחברות יש מספר נקודות להתייחסות- פיצול חברה לישראל.
* אורמת נקנתה על ידי חברת הבת שלה אורמת טכנולגיות.
* פריגו ישראל.

דגשים נוספים לחיבור הנתונים:

* לוודא דיוודנדים.
* לוודא ימי מסחר פעילים.
* אינדוקס יומי ויצירת time series

הפלט:

* יוציא טבלת RData אשר בה :
  + העמודות יהיו החברות. symbol
  + השורות יהיו התשואה היומית r / relative price
  + אינדקסים של time series (סדרה עיתית), בתדירות יומית (ככל הנראה המימד השלישי בטבלה)
  + StockPrices - (price,relative\_p,open/close)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | ....... | S2open |  | S1return | S1price | S1close | S1open | Date |
|  |  |  |  |  |  |  |  |  |  | 1.1.2006 |
|  |  |  |  |  |  |  |  |  |  | 31.07.2016 |

\*חישוב r זה בין Open ל Close של אותו יום.

\*המחירים מתואמים, אין צורך להתחשב בדיווידנדים.

1. פונקציה שניה K\_Histogram מקבל את StockPrices ומספר הימים K. בנוסף מקבלת תאריך אחרון לחישוב ומיתר התאריכים מתעלמת. התאריך האחרון בשם DDate. ושומרת על

אם יש מידע קיים , והתאריך מאוחר יותר אזי יש להוסיף רק את התאריכים הרלוונטים.

אם נכנס תאריך מוקדם מהמידע השמור צריך לעשות reset לכל החישוב.

פונקציה זו תקבל את מבנה הנתונים המתואר בפונקציה Stock Data.

תפקיד הפונקציה – פונקציה זו תקצוב את רצף הזמן לתקופות מסחר (מסחר יומי, דו יומי, שבועי וכו..). כל תקופה נעה בחלון , כל יום החלון זז צעד זמן אחד (יש לשים לב להערה 2). (ללא חופשות וחגים – #חג זה מעל ליומיים ).

דגשים:

* יש לתת חשיבות לתקופות של מסחר, אני לא בטוח שזה נכון לעשות צמדים של יום ברגע שיש חגים באמצע או שבת. לצורך העניין ככל הנראה שלא יהיה קשר בין יום ה' ליום א' בהינתן ואנו עושים תקופה של יומיים מאחר ובין חמישי לראשון יש הרבה רעש.. נקודה למחשבה.
* ברגע שיש חגים הקוטעים את המסחר אזי יש לעצור את החלון הנע,. הגדרה כל הפסקת מסחר מעל יומיים נתייחס לזה כחג. עם זאת יש להמשיך בתקופה שזה ממשיך ישירות לאחר מכן# בתקופת החג יש ליצור חלון קבוע באותו גודל (אולי נשמיט את זה בסוף לרשום הערות בנושא זה(

פלט:

וקטור של אינדקס, האורך שלו יהיה קטן מ K-N (כאשר N זה מספר ימי התקופות)

כל אינדקס הוא נקודת ההתחלה.

|  |
| --- |
| KVec |
| 1 |
| 6 |
| 16 |
| .... |

1. פונקציה שלישית- Classification מקבלת KVec (אינדקסים) StockPrices L- פרמטר למספר סוגי הימים.

פונקציה זו תקבל את המידע מ K-Histogram, ותקבל פרמטר L אשר אומר איך לסווג את התקופות מסחר.

* הפונקציה תשתמש בפעם הראשונה במודל K-means.
* הפונקציה תשתמש בפעם השנייה עם kernel, עוד לא נכנסו לעובי הקורה בנושא.
* דיברנו גם על tree ועל random forest.

פלט: הפונקציה תוציא מספר רשימות שונות:

1.תוציא את הסיווג ה Li לכל תקופה. ה KVec ועם עמודה נוספת של סיווג ששם העמודה הוא המסווג:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Monkey | RandomForest | Kernel | K\_Means | KVec |
|  |  | 1 | 2 | 1 |
|  |  | 2 | 3 | 6 |
|  |  | 7 | 4 | 16 |
|  |  | 1 | 3 | .... |

Min == 1; Max == L;

שם הקובץ.Classifier.RData