

بسمه تعالی

داده کاوی- پیش بینی قیمت ده ارز و فلز گرانبها

▪ مقدمه

همواره اقشار مختلف جامعه از اقتصاددانان گرفته تا عامه مردم به بازار ارز و فلزات گرانبها توجه ویژه ای داشته اند. در این زمینه، یکی از دغدغه های همیشگی اهل فن دستیابی به یک مجموعه پیش بینی نسبتا دقیق از قیمت (نرخ) انواع ارز و فلزات گرانبها بوده است. مساله پیش رو، طراحی و بکارگیری روشی برای مواجهه با چالش مذکور است.

▪ تعریف مساله

زمان در اقتصاد نقش مهمی دارد. گواهی بر این مدعا قیمت هر اونس طلا است که اگر مورد پرسش قرار بگیرد، سوال دیگری هم در کنارش مطرح است و آن این است که در چه زمانی؟ از طرف دیگر، قیمت طلا و دیگر فلزات گرانبها و حتی قیمت ارزها در طول زمان مقداری تصادفی نبوده و روندی پیوسته در امتداد تاریخچه آن دارد. بنابراین، اگر این روندهای تغییر قیمت به خوبی شناسایی و درک شوند، می توان پیش بینی نزدیک به واقعیتی از میزان آنها در آینده داشت. هدف از برگزاری این مسابقه، طراحی و بکارگیری مدلی برای پیش بینی دقیق قیمت ده قلم کالا (ترکیبی ارز و فلز گرانبها) بر طبق قیمت آنها در طول زمان است.

برای این منظور، مجموعه داده ای در اختیارتان قرار می گیرد. این مجموعه داده، حاوی اطلاعاتی است که هر پنج دقیقه جمع آوری شده بدین ترتیب که در هر بار، ۱۲۰ رکورد آخر مربوط به قیمت های جدید ده قلم هدف و آمارهای دیگر ثبت و ذخیره شده است. فیلدهای مختلف این مجموعه داده عبارتند از:

فرمت داده:

داده ای که در اختیار شما قرار گرفته (این داده ها در پرتال رقابت به آدرس contest.fanavard.com قابل دسترسی است) برای ۹ ارز/فلز گران بها از یکی از بازار های دنیا جمع آوری شده است. برای هر کالا ۳ فایل در اختیار شما قرار گرفته است که به

ترتیب مربوط به خلاصه وضعیت بازار، معاملات صورت گرفته و سفارش ها ثبت شده است.

فایل مربوط به خلاصه بازار:

اسم این فایل برای کالای A به صورت A_ticker است. ستون های این فایل به ترتیب شامل موارد زیر است:

۱. زمان ثبت اطلاعات که با رزولوشن ۵ دقیقه جمع آوری شده است
 ۲. قیمت حداکثر سفارش ثبت شده برای فروش
 ۳. حجم حداکثر سفارش ثبت شده برای فروش
 ۴. قیمت حداکثر سفارش ثبت شده برای خرید
 ۵. حجم حداکثر سفارش ثبت شده برای خرید
 ۶. قیمت آخرین معامله صورت گرفته در بازار
 ۷. حجم معاملات روزانه
 ۸. حداکثر قیمت روزانه
 ۹. حداقل قیمت روزانه
- فایل مربوط به معاملات صورت گرفته:

اسم این فایل برای کالای A به صورت A_trades است. ستون های این فایل به ترتیب شامل موارد زیر است:

ستون اول زمان ثبت اطلاعات که با رزولوشن ۵ دقیقه جمع آوری شده است
هر معامله با ۴ عدد مشخص می شود. Id منحصر به فرد برای هر معامله. زمان معامله به میلی ثانیه. حجم معامله و قیمت معامله. برای هر لحظه مشخصات ۱۲۰ معامله ثبت شده است. لذا ادامه فایل شامل 120×4 ستون است که هر ۴ ستون کنار هم به ترتیب مربوط به یک معامله هستند. و به ترتیب ذکر شده می باشند.

فایل مربوط به سفارش ها ثبت شده:

اسم این فایل برای کالای A به صورت A_book است. ستون های این فایل به ترتیب شامل موارد زیر است:

ستون اول زمان ثبت اطلاعات که با رزولوشن ۵ دقیقه جمع آوری شده است

سفارش ها ثبت شده به صورت histogram از وضعیت بازار ارائه می شوند. و کل این histogram به صورت ۵۰ bin (۲۵ تا بالاتر و ۲۵ تا پایین تر از قیمت اصلی) ارائه می شوند. هر bin به ترتیب شامل قیمت تعداد سفارش های ثبت شده مربوط به این قیمت و حجم سفارش ثبت شده مربوط به این قیمت می باشد. بنابراین این فایل در ادامه شامل ۳ × ۵۰ ستون است که هر ۳ ستون کنار هم به ترتیب مربوط به یک bin از سفارش های ثبت شده هستند. و به ترتیب ذکر شده می باشند.

* نکته: هنگامی که به هر دلیلی اطلاعات ناصحیحی برای اطلاعات فوق جمع آوری شود، در آن سطر داده مقدار ۱- درج شده است.

▪ ارزش ها

یک مدل «قابل قبول و مناسب» را موارد زیر از یک مدل «غیرقابل قبول» متمایز می کند:

- **مبتنی بر درک صحیح و کاملی از مساله:** نیاز است نشان داده شود که مساله به خوبی درک شده و مدل، برآمده از درک مساله است و نه صرفا نتیجه چندین آزمایش بر روی روش های مختلف.
- **سادگی در اجرا:** یک مدل قابل قبول باید بدون نیاز به نصب نرم افزار یا وجود سخت افزار خاص روی رایانه ای معمولی با سیستم عامل ویندوز قابل استفاده و آزمایش باشد.
- **اعتبار علمی:** مدلی که پشتوانه علمی معتبر نداشته باشد و اصطلاحا من درآوردی باشد قابل قبول نیست.

- **مستندسازی کامل:** نیاز است تمامی مراحل طی شده، روش ها و استدلال ها به خوبی توضیح داده شود.
- **بصری سازی:** توضیحات بدون نمودار قابل قبول نخواهد بود.

▪ نحوه شرکت در مسابقه

لازم است شرکت کنندگان فایل اجرایی مدل خود را به همراه توضیحات (در یک فایل PDF) را ارسال کنند.

▪ نحوه سنجش

موارد زیر در انتها سنجیده خواهد شد. هر بخش، نمره‌ای از ۱۰۰ خواهد داشت.

- ✓ میزان درک مساله
- ✓ اجرای راحت و سریع فایل اجرایی مدل
- ✓ کیفیت فایل توضیحات
- ✓ کیفیت نتایج مدل

▪ زمان بندی و شرایط ارزیابی:

پایان مرحله اینترنتی برای هر شرکت کننده تاریخ ۲۹ اردیبهشت ماه ۹۷ پس از ثبت نام و دریافت فایل توضیحات مسابقه می باشد.

فهرست اسامی شرکت کنندگان راه یافته به مسابقه نهایی از طریق ایمیل و پیامک به ایشان اطلاع رسانی خواهد شد.

در بخش حضوری، افراد راه یافته به مرحله دوم فرصت خواهند داشت در رقابتی فشرده تر طرح خود را تکمیل کرده و یا بهبود بخشند، لازم به ذکر است الزامی از جانب برگزارکننده جهت بهبود و تکمیل طرح ها وجود نخواهد داشت بلکه این مرحله صرفا فرصت مجددی است برای شرکت کنندگانی که خود را در یک قدمی برنده شدن می بینند!

▪ **جوایز:**

فن‌آورد به جهت تقدیر از شرکت کنندگان و برگزیدگان مسابقه جوایزی را ترتیب داده است که بدین شرح است:

- جایزه نفر اول ۱۵ میلیون تومان + لوح تقدیر + گواهینامه معتبر مهارتی + رتبه‌بندی ملی تخصصی
- جایزه نفر دوم ۵ میلیون تومان + لوح تقدیر + گواهینامه معتبر مهارتی + رتبه‌بندی ملی تخصصی
- جایزه نفر سوم ۲ میلیون تومان + لوح تقدیر + گواهینامه معتبر مهارتی + رتبه‌بندی ملی تخصصی
- لوح تقدیر به نفراتی که حداقل معیار قبولی طرح را اخذ کرده باشند.
- گواهی حضور در مسابقات فن‌آورد به افرادی که کف نمره داوری را اخذ کنند.

▪ **پاسخگویی:**

همچنین شرکت کنندگان می‌توانند پرسش‌های خود را پیرامون این رقابت از طریق آدرس ایمیل زیر با دبیرخانه فن‌آورد در مسابقه "داده‌کاوی" فن‌آورد طرح نمایند.

data@fanavard.com

منتظر شما هستیم.