

۱- مقدمه

امروزه با رشد و توسعه شبکه های اجتماعی، این شبکه ها نقش مهمی در زندگی روزمره به خود اختصاص داده اند. اقشار مختلف اجتماع جهت بهره گیری از امور مختلفی از جمله آموزش، کسب و کار، سرگرمی و ... به این شبکه ها مراجعه می کنند. از این رو تحلیل داده هایی که در این شبکه ها در حال انتشار است از اهمیت ویژه ای برخوردار بوده و می تواند در بسیاری از موارد مفید باشد. هدفی که در این مسابقه دنبال می شود بکارگیری و تحلیل این داده ها برای پیش بینی نتایج انتخابات پیش رو در حوزه ی انتخابیه ی تهران است. شایان ذکر است که به گروه های اول تا سوم این مسابقه جوایز نفیسی اهدا خواهد شد.

۲- تعریف مساله

هدفی که در این مسابقه دنبال می شود پیش بینی نتایج انتخابات یازدهمین دوره مجلس شورای اسلامی در حوزه ی انتخابیه ی تهران است. در این انتخابات، تهران دارای سی کرسی بوده که انتظار می رود شرکت کنندگان، نمایندگان این سی کرسی را به ترتیب پیش بینی کنند.

برای انجام این کار انتظار می رود شرکت کنندگان الگوریتم یا مدلی را ارائه داده و با تحلیل داده های منتشر شده در شبکه های اجتماعی، افرادی را که شانس موفقیت بیشتری در انتخابات دارند را، شناسایی کنند. خروجی مورد انتظار لیستی شامل ۳۰ کاندید بوده که با بیشترین احتمال در انتخابات یازدهمین دوره مجلس شورای اسلامی، رای خواهند آورد. در کنار آن گزارش کتبی نحوه ی کارکرد مدل و داده هایی که برای پیش بینی مورد استفاده قرار گرفته است نیز بایستی ارائه گردد. نحوه ی ثبت نام و شرکت در مسابقه، در [سایت رسمی مسابقه](#) آورده شده است.

جمع آوری داده های شبکه های اجتماعی در مسابقه برعهده ی تیم های شرکت کننده بوده و استفاده از هر مجموعه داده ای برای آموزش مدل ها، بلامانع است. در کنار آن داده های مربوط به تعداد محدودی از کانال های تلگرامی نیز، در اختیار شرکت کنندگان قرار خواهد گرفت.

۳- قوانین شرکت در مسابقه

- ارائه مشخصات مسئول تیم در زمان ثبت نام الزامی است.
- کلیه مکاتبات با مسئول تیم انجام خواهد شد و در صورت کسب مقام، جایزه به مسئول تیم تحویل داده خواهد شد.
- هر فرد تنها می تواند عضو یک تیم شرکت کننده در مسابقه باشد.
- محدودیت سنی برای شرکت در مسابقه وجود ندارد.
- محدودیتی در تعداد تیم های شرکت کننده از سوی یک سازمان یا نهاد خاص وجود ندارد.
- ارائه گزارش فنی، سورس کد و داده های مورد استفاده برای ارزیابی تیم ها الزامی است.
- استفاده از کلیه داده های شبکه های اجتماعی مانند تلگرام، توییت و ... آزاد است (جمع آوری این داده ها برعهده تیم های شرکت کننده می باشد). با این حال برای نمونه داده های مربوط به چند کانال تلگرامی پس از اعلام نهایی اسامی کاندیدا در اختیار شرکت کنندگان قرار خواهد گرفت. این داده ها از کانال های عمومی تلگرام گردآوری خواهد شد. **بدیهی است این داده ها صرفا جنبه راهنمایی داشته و اجباری به استفاده از آنها نیست** (این مجموعه هیچ مسئولیتی در مورد محتوای پیام های منتشر شده از شبکه های اجتماعی به عهده نمی گیرد).

۴- نحوه ارزیابی تیم ها

به منظور ارزیابی تیم ها دو معیار اصلی مورد استفاده قرار می گیرد. معیار اول براساس درصد صحت پیش بینی تیم ها از نتایج انتخابات محاسبه گردیده و معیار دوم امتیازی است که به گزارش فنی تیم ها اختصاص می یابد. در گزارش فنی هر تیم بایستی به معرفی مدل و داده هایی که از آنها برای پیش بینی استفاده شده است پرداخته شود. مجموع این دو معیار، امتیاز نهایی و رتبه بندی تیم ها را مشخص خواهد کرد.

۵- چند پیشنهاد برای پیش‌بینی نتایج انتخابات

در این بخش به معرفی چند ایده و الگوریتم پرداخته می‌شود که شرکت‌کنندگان می‌توانند در مدل‌های پیشنهادی خود از آن‌ها استفاده کنند. این بخش تنها جنبه راهنمایی داشته و اجباری در بکارگیری این روش‌ها وجود ندارد. شرکت‌کنندگان می‌توانند از هر روش و ایده‌ای که در حوزه یادگیری ماشین و پردازش زبان‌های طبیعی قرار می‌گیرد، استفاده نمایند. از ایده‌ها و روش‌های جدید استقبال می‌شود.

یکی از کارهایی که برای پیش‌بینی می‌توان انجام داد این است که به هر کاندیدا وزنی نسبت داده شود که در ابتدا مقدار صفر دارد. در ادامه با تجزیه و تحلیل متن هر پیام، وزن کاندیداهایی که راجع به آن‌ها صحبت شده است، تغییر کند. این تغییر می‌تواند بسته به محتوای پیام مثبت یا منفی باشد. تغییر وزن نیز می‌تواند براساس یکی از سیاست‌هایی که در ادامه آورده شده است، انجام شود.

- شمارش تعداد برخورد:

وزن هر پیام را می‌توان برابر با ۱ در نظر گرفت و در صورت مرتبط بودن با هر کدام از کاندیداها برای آن یک امتیاز در نظر گرفته می‌شود. نتیجه نهایی با شمارش تعداد پیام‌های مربوط به هر کاندیدا و سپس درصدگرفتن محاسبه می‌گردد.

- در نظر گرفتن اعتبار منتشرکننده‌ی پیام:

اعتبار و تاثیرگذاری پیام‌ها بسته به منبع انتشارشان متفاوت خواهد بود. برای مثال پیامی که در یک کانال تلگرامی منتشر می‌شود بسته به تعداد اعضای آن کانال و میزان تخصص آن کانال در موضوعات انتخاباتی، تاثیرگذاری متفاوتی دارد. با در نظر گرفتن این عوامل، می‌توان به پیام‌های متفاوت وزن‌های متفاوتی داد و در نهایت به دقت بهتری در پیش‌بینی دست یافت.

- تحلیل محتوایی پیام‌ها:

تحلیل محتوایی که توسط پیام منتقل می‌شود از اهمیت ویژه‌ای برخوردار است. به‌عنوان مثال می‌توان برای هر پیام یک یا چند موجودیت هدف مشخص کرد و سپس به بررسی این موضوع پرداخت که برای هر کدام از موجودیت‌ها چه احساسی منتقل شده است و از آن برای امتیازدهی به کاندیداها استفاده کرد.

در تهران مرسوم است که تعدادی از کاندیداها گروه‌هایی را تشکیل می‌دهند. در بعضی از پست‌های شبکه‌های اجتماعی نیز راجع به آن گروه صحبت می‌شود. در این‌گونه موارد می‌توان امتیاز کل افرادی که در آن گروه خاص قرار می‌گیرند را براساس یکی از روش‌های فوق تغییر داد.

۶- معرفی چند منبع که می‌تواند مفید باشد

- [پارسی‌ور: ابزار پیش‌پردازش متون فارسی](#)
- [مجموعه صفات دارای برچسب قطبیت \(لکسی پرس\)](#)

۷- تماس با ما

جهت کسب آخرین اطلاعات و اخبار مرتبط با مسابقه در کانال تلگرامی مسابقات عضو شده و وب سایت مسابقات را مطالعه فرمایید.

سوالات، انتقادات و پیشنهادهای خود در مورد مسابقات را نیز می‌توانید از طریق ایمیل یا تلفن با ما در میان بگذارید.

Website: PrediComp.ictrc.ac.ir

Email: PrediComp@ictrc.ac.ir

Telegram: <https://t.me/predicomp>

تلفن: ۰۲۱-۸۸۹۳۰۱۵۰ داخلی ۴۱۱