شرکت شبکه الکترونیکی پرداخت کارت (سهامی خاص)
شاپرک

9th Annual Conference on
Electronic Banking & Payment Systems

پژوهشکده پولی و بانکی
بانک مرکزی جمهوری اسلامی ایران

بانک مرکزی جمهوری اسلامی ایران

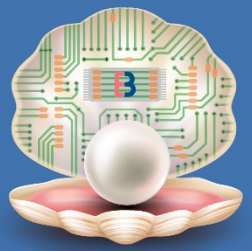
شرکت ملی انفورماتیک

نهمین همایش سالانه
بانکداری الکترونیک
و نظام‌های پرداخت

(ارزش‌آفرینی دیجیتال)

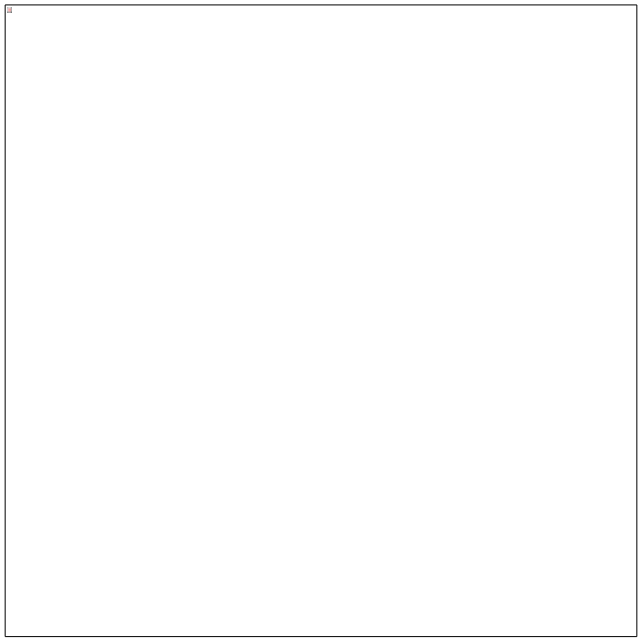
تهران، مرکز همایش‌های بین‌المللی برج میلاد - ۱ و ۲ اسفند ۱۴۰۱

ترموگراف: ترکیب روش‌های متن‌کاوی و گراف‌کاوی به‌منظور پیش‌بینی اصناف در شبکه‌ی پرداخت



ارائه‌دهنده: بهناز پورابراهیم

مقدمه



۱ شبکه پرداخت الکترونیک

۲ پذیرنده و کد صنف مرتبط با فعالیت آن

ناهنجاری صنف

تعریف

عدم ثبت کد صنف مرتبط با فعالیت پذیرندگان

هدف

تشخیص ناهنجاری اصناف

چالش‌ها

وجود ناهنجاری صنف و عدم تشخیص آن برای پذیرندگان

هدف

۱ تجزیه و تحلیل نام پذیرندگان

۲ تشخیص نوع فعالیت صنفی هر پذیرنده

۳ پیش‌بینی صنف درست جایگزین

۴ تکنیک‌های متن‌کاوی و گراف‌کاوی

مجموعه‌داده

- مجموعه S شامل صنف‌های ۱ تا N هستند.

$$S = \{s_1, s_2, \ldots, s_N\}$$

- هر صنف s_i شامل k پذیرنده (p) است.

$$s_i = \{p_1, p_2, \ldots, p_k\}$$

- مجموعه‌داده‌ی ما مجموعه نام فروشگاهی و شامل هشت میلیون پذیرنده است.

$$Dataset = \{p_1, p_2, \ldots, p_{8000000}\}$$

مدل پیشنهادی: ترموگراف

قدم اول

تحلیل متن‌کاوی نام پذیرندگان و محاسبه‌ی شباهت معنایی بین اصناف

قدم دوم

ساخت گراف میان اصناف باتوجه به شباهت معنایی اصناف

قدم سوم

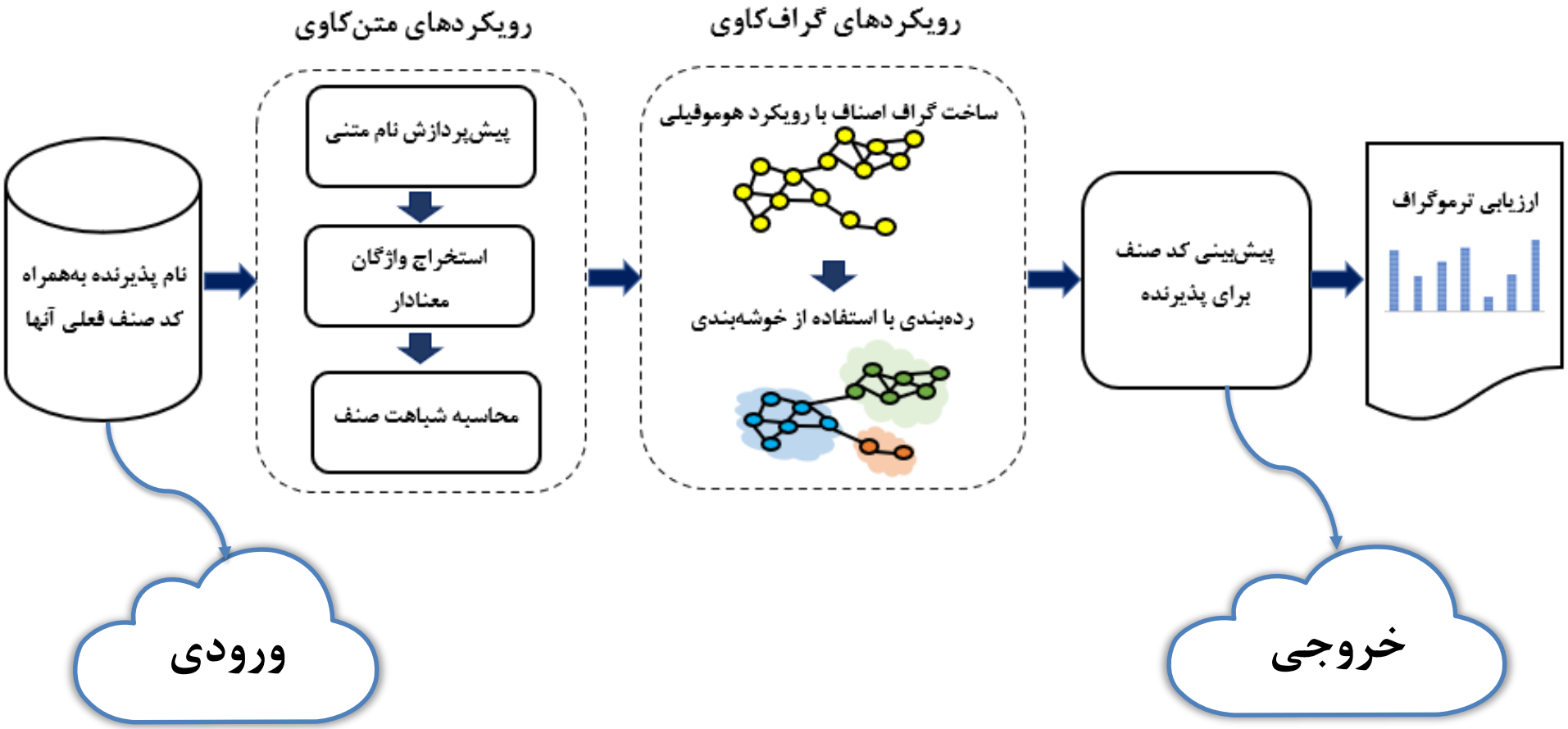
پیش‌بینی کد صنف پذیرندگان با رویکردهای گراف‌کاوی و خوشه‌بندی

نوآوری‌ها

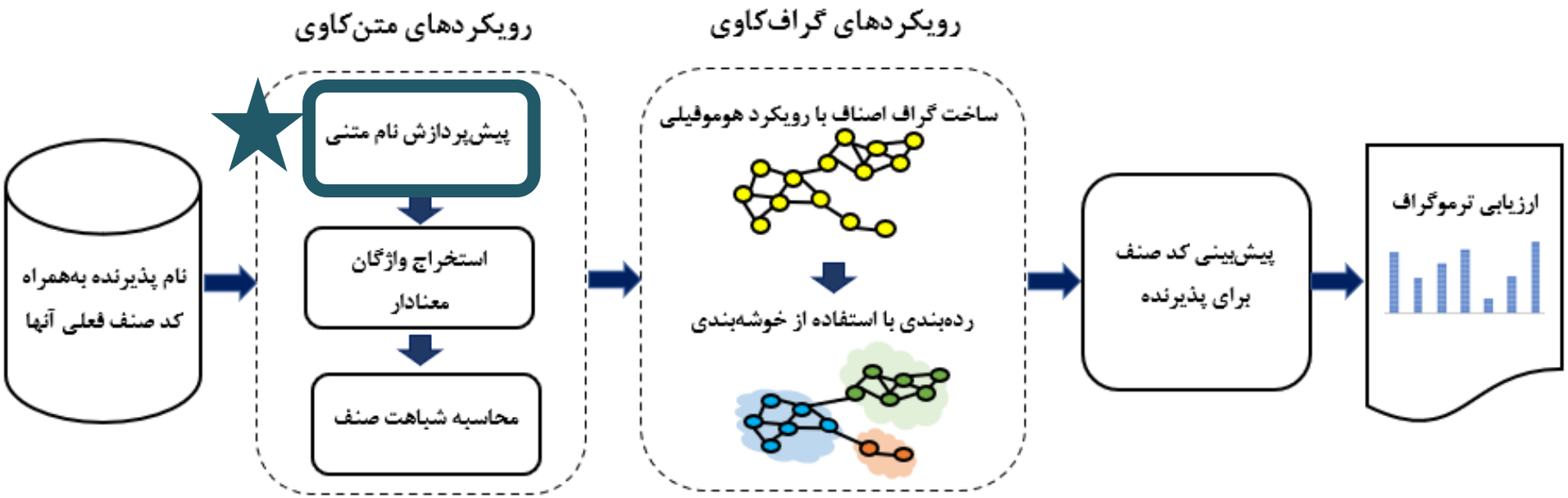
الف) استفاده از داده‌های متنی نام پذیرندگان

ب) رویکرد ترکیبی جدید متن‌کاوی و گراف‌کاوی

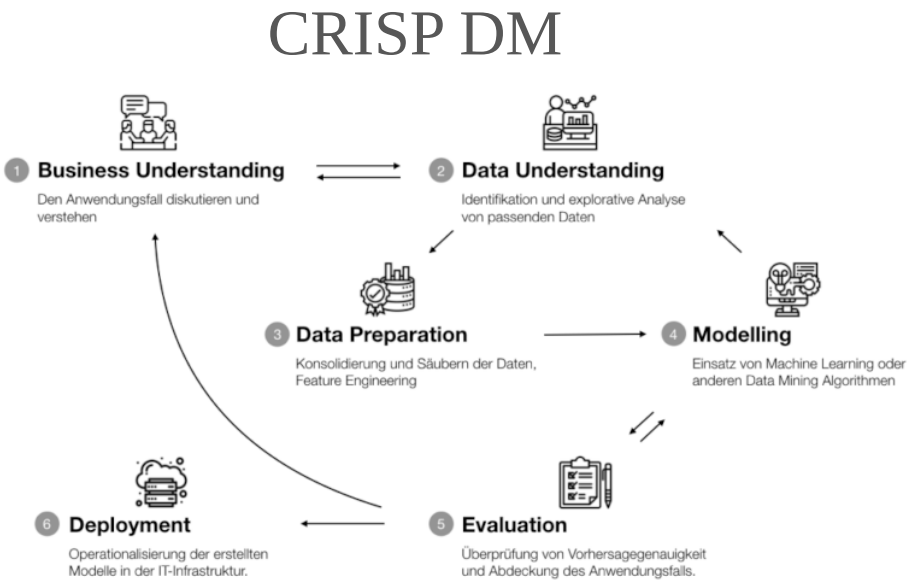
معماری کلی ترموگراف



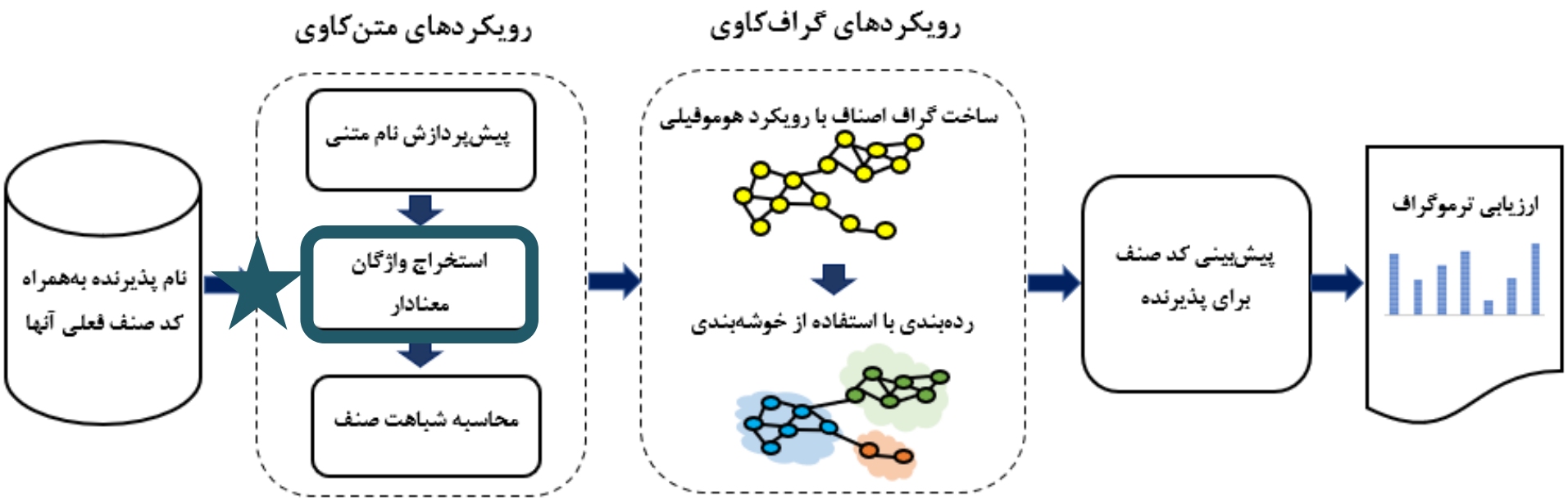
معماری کلی ترموگراف



پیش‌پردازش داده‌های متنی



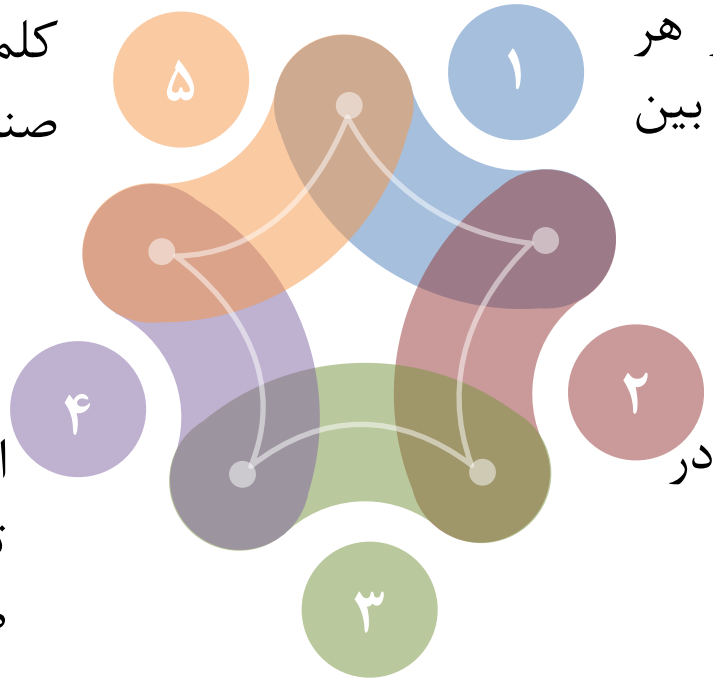
معماری کلی ترموگراف



استخراج واژگان معنادار

نام پذیرنده‌های موجود در هر صنف براساس فاصله‌ی بین کلمات استخراج می‌کنیم.

کلمات معنادار و مرتبط با صنف را استخراج می‌کنیم.



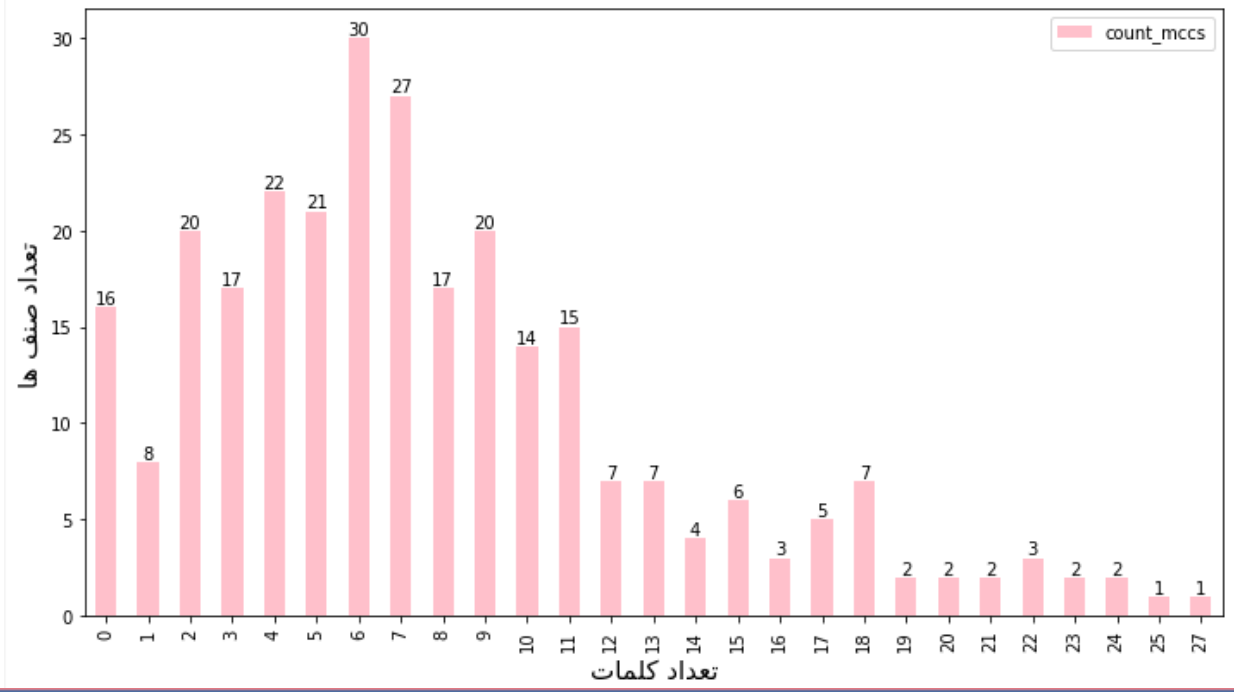
تعداد تکرار هر کلمه در صنف نیز محاسبه می‌شود.

از بین ۱۰۰ کلمه‌ی کاندید، تعدادی کلمه مرتبط معنادار با صنف و تعدادی کلمه نامرتبط با صنف وجود دارد.

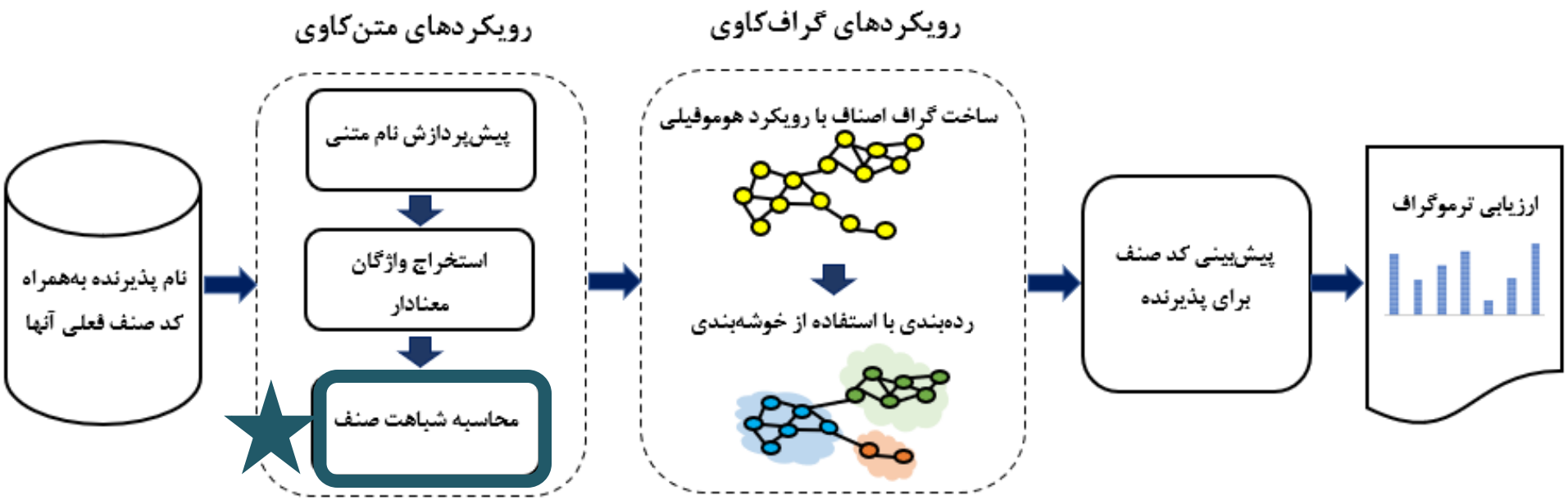
۱۰۰ کلمه‌ی پرتکرار هر صنف به‌عنوان کلمات کاندید صنف استخراج می‌شوند.

تنوع و پراکندگی نام پذیرندگان

- برای بعضی از صنوف تعداد کلمات معنادار **زیادی** انتخاب شده است که دلیل آن **تنوع** نام پذیرنده‌ها است.
- برای بعضی از صنوف تعداد کلمات معنادار **کمی** انتخاب شده است که دلیل آن **پراکندگی** نام پذیرنده‌ها است.



معماری کلی ترموگراف



محاسبه‌ی شباهت اصناف

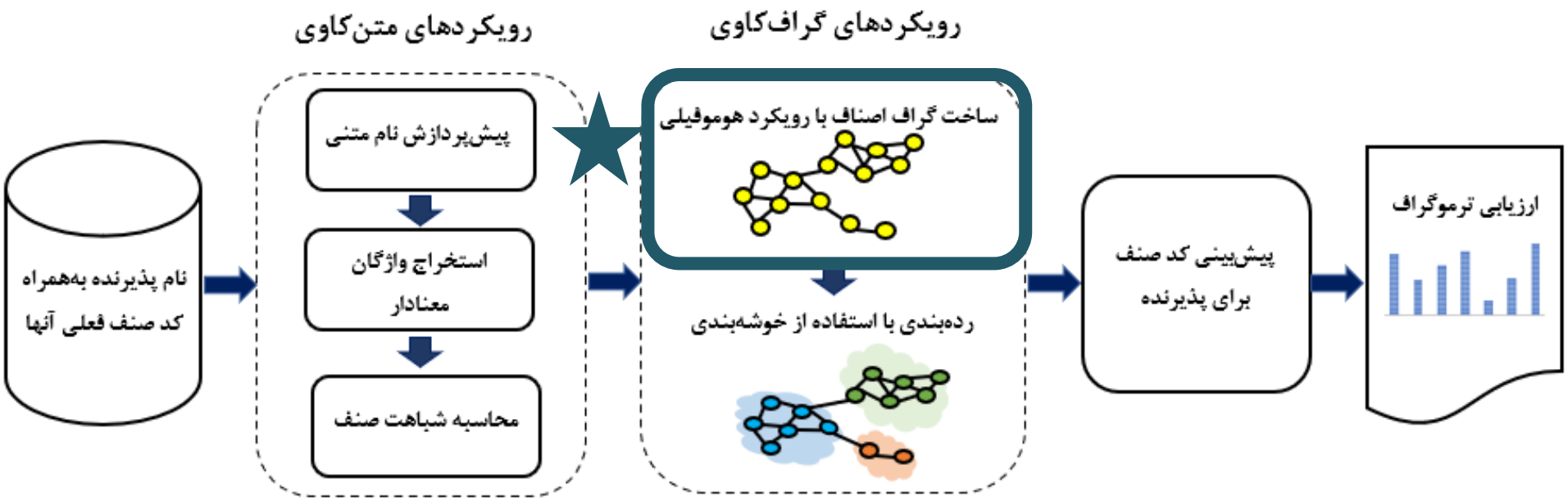
۱ امکان دارد یک کلمه در چند صنف مشترک باشد.

۲ صنف‌هایی که با هم کلمه مشترک دارند مشابه در نظر می‌گیریم.

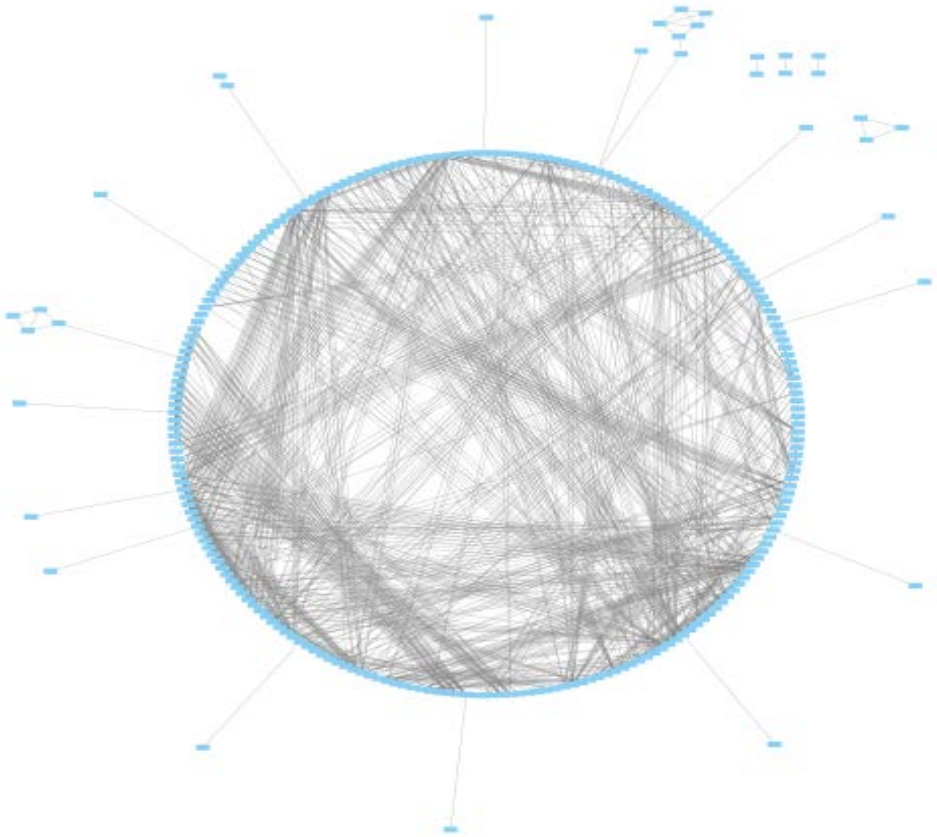
۳ هرصنف با تعدادی صنف دیگر مشترک است.

۴ شباهت دو صنف به تعداد کلمات معنادار مشترک بین دو صنف بستگی دارد.

معماری کلی ترموگراف



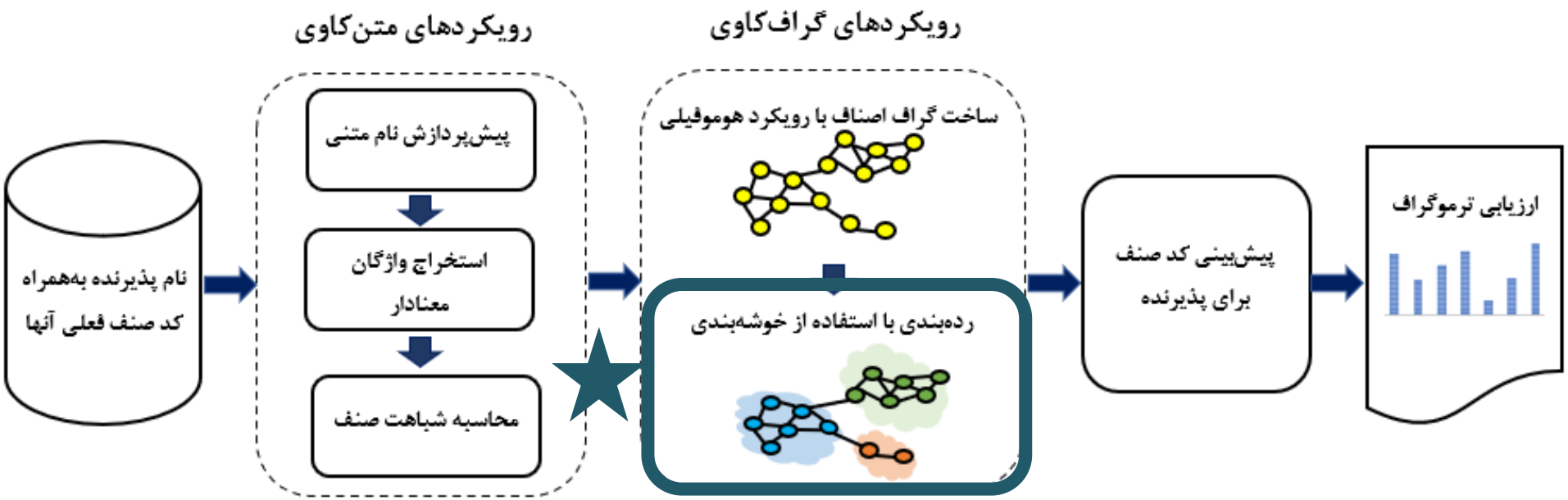
تحلیل گرافی اصناف



- گره: صنف
- یال: کلمه معنادار مشترک

- تحلیل گرافی اصناف: اگر بین هر دو صنف حداقل یک کلمه‌ی مشترک وجود داشته باشد، دو صنف با یک یال به‌هم متصل می‌شوند.

معماری کلی ترموگراف



خوشه‌بندی اصناف

چالش تحلیل ریزدانه

عدم وجود برچسب

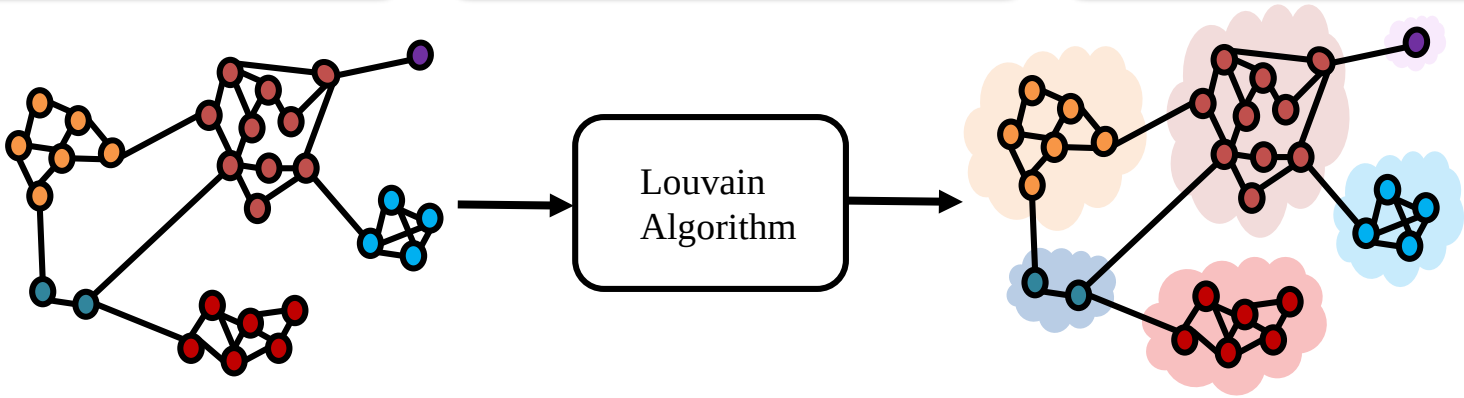
نیاز به تحلیل درشت دانه

خوشه بندی (تحلیل بدون ناظر)

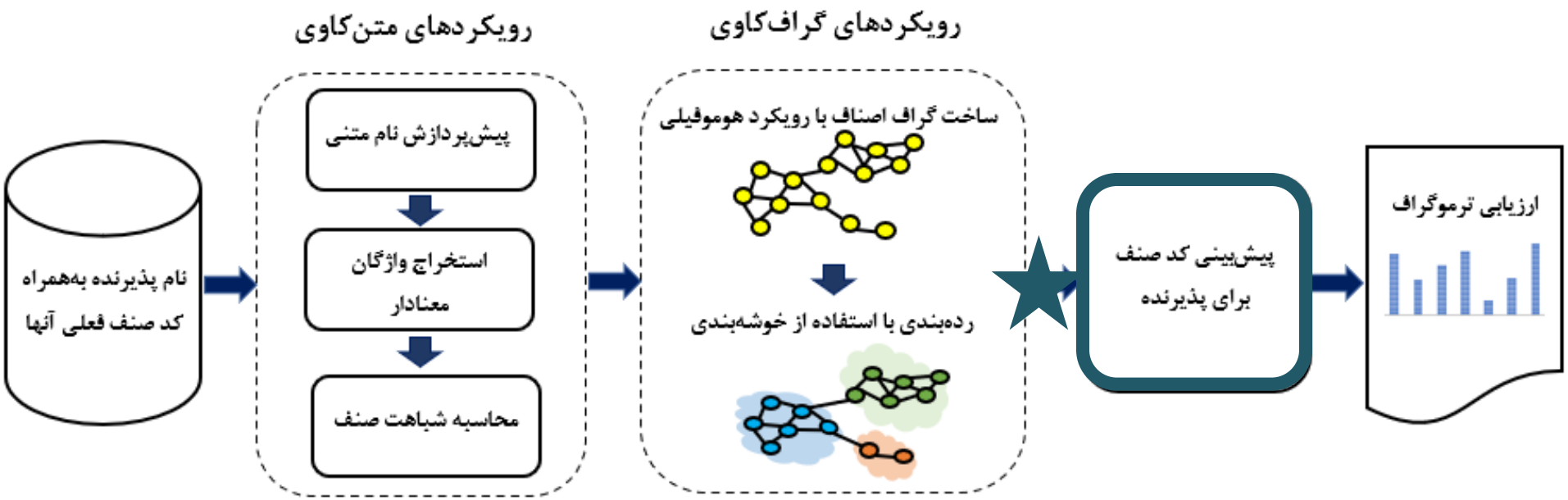
خوشه بندی گرافی Louvain

شباهت بین خوشه ای پایین

شباهت درون خوشه ای بالا



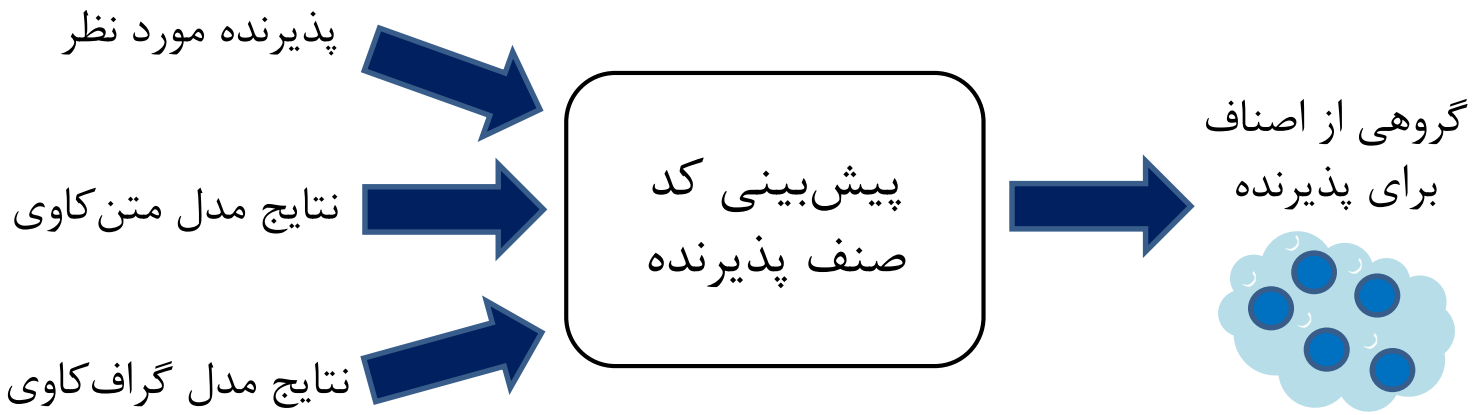
معماری کلی ترموگراف



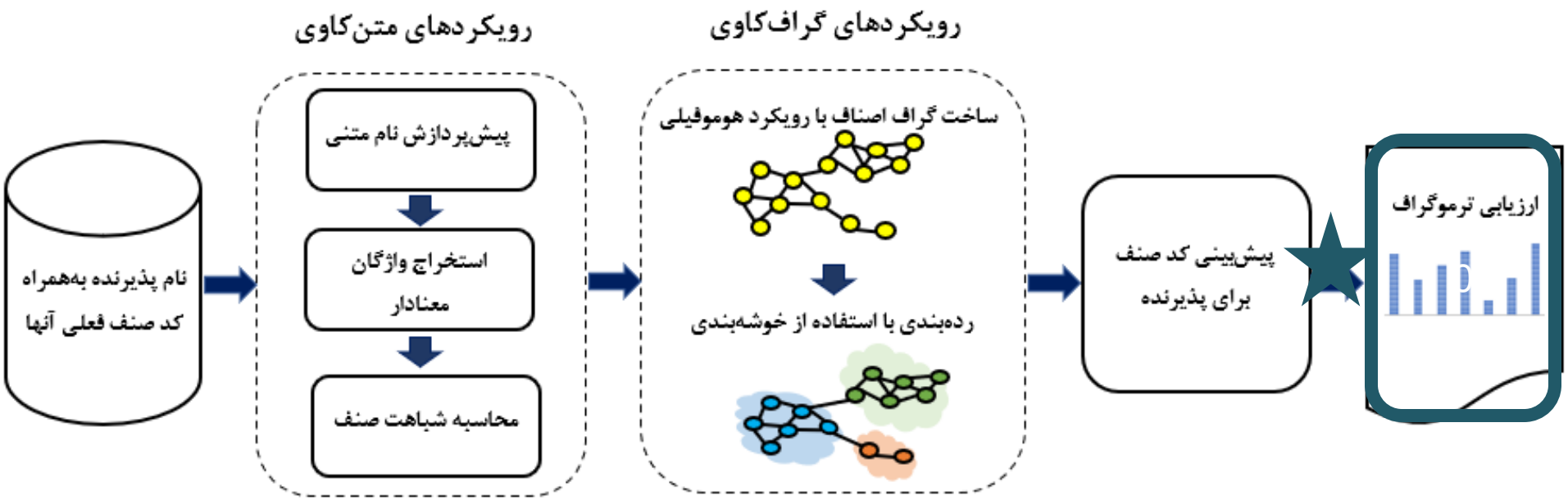
فیلترهای اعمال شده روی نام پذیرنده

پیش‌بینی کد صنف پذیرنده

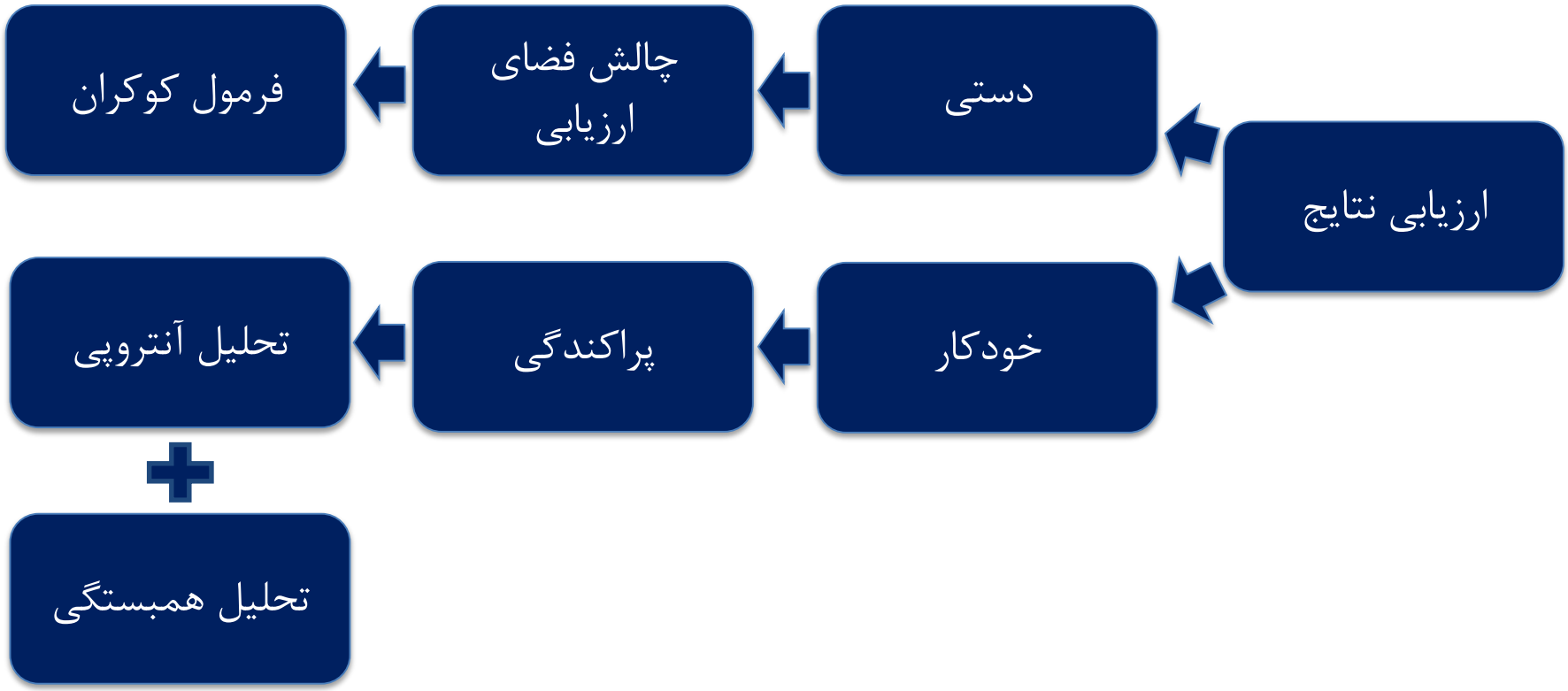
- کلمات معنادار هر صنف را به لیست تبدیل می‌کنیم.
- اگر از این کلمات در نام پذیرنده‌ای وجود داشته باشد، درنتیجه صنف فعلی پذیرنده اشتباه ثبت شده.
- صنف موردنظر و خوشه‌ای که صنف در آن قرار دارد برای آن پیش‌بینی می‌شود.



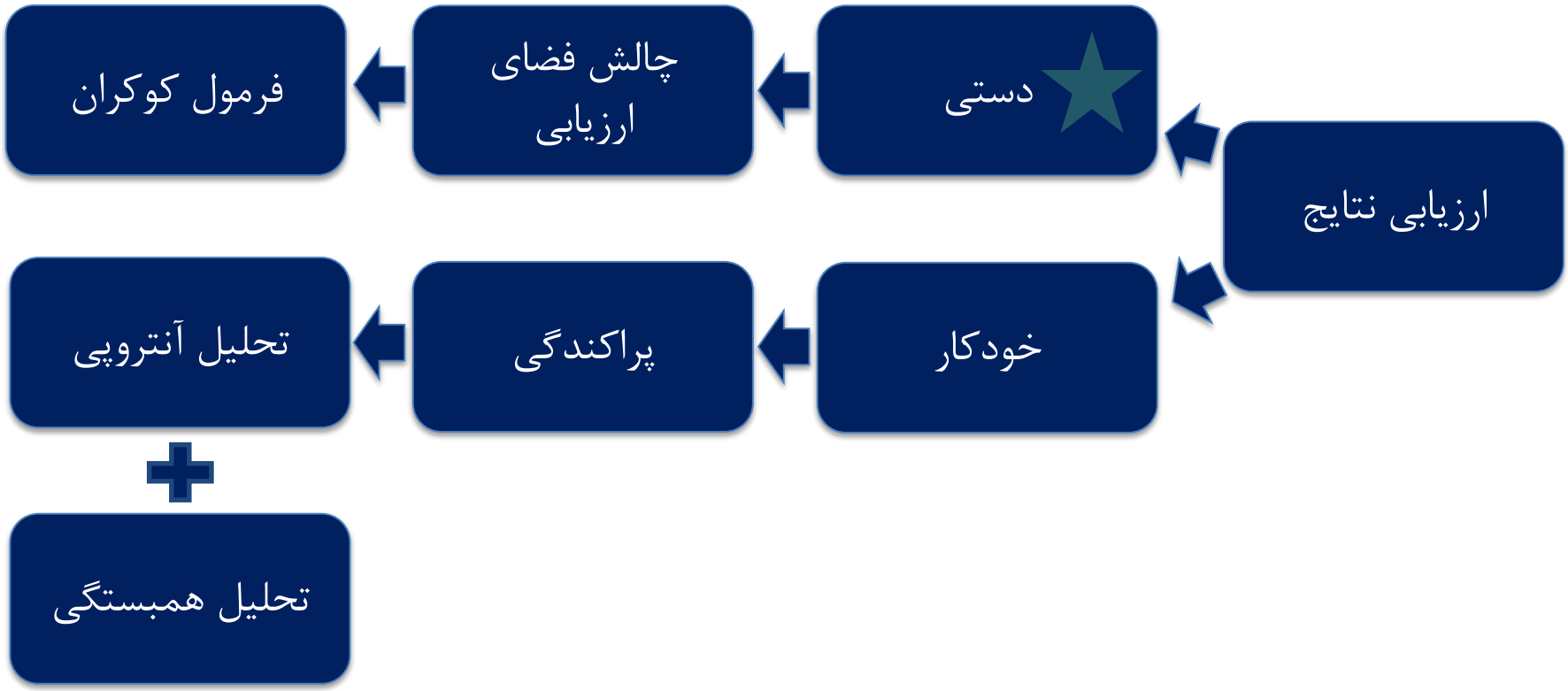
معماری کلی ترموگراف



ارزیابی نتایج

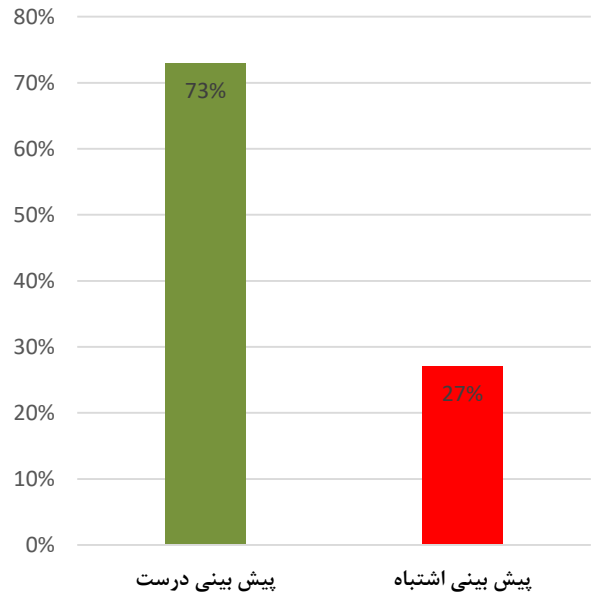


ارزیابی نتایج



ارزیابی دستی نتایج

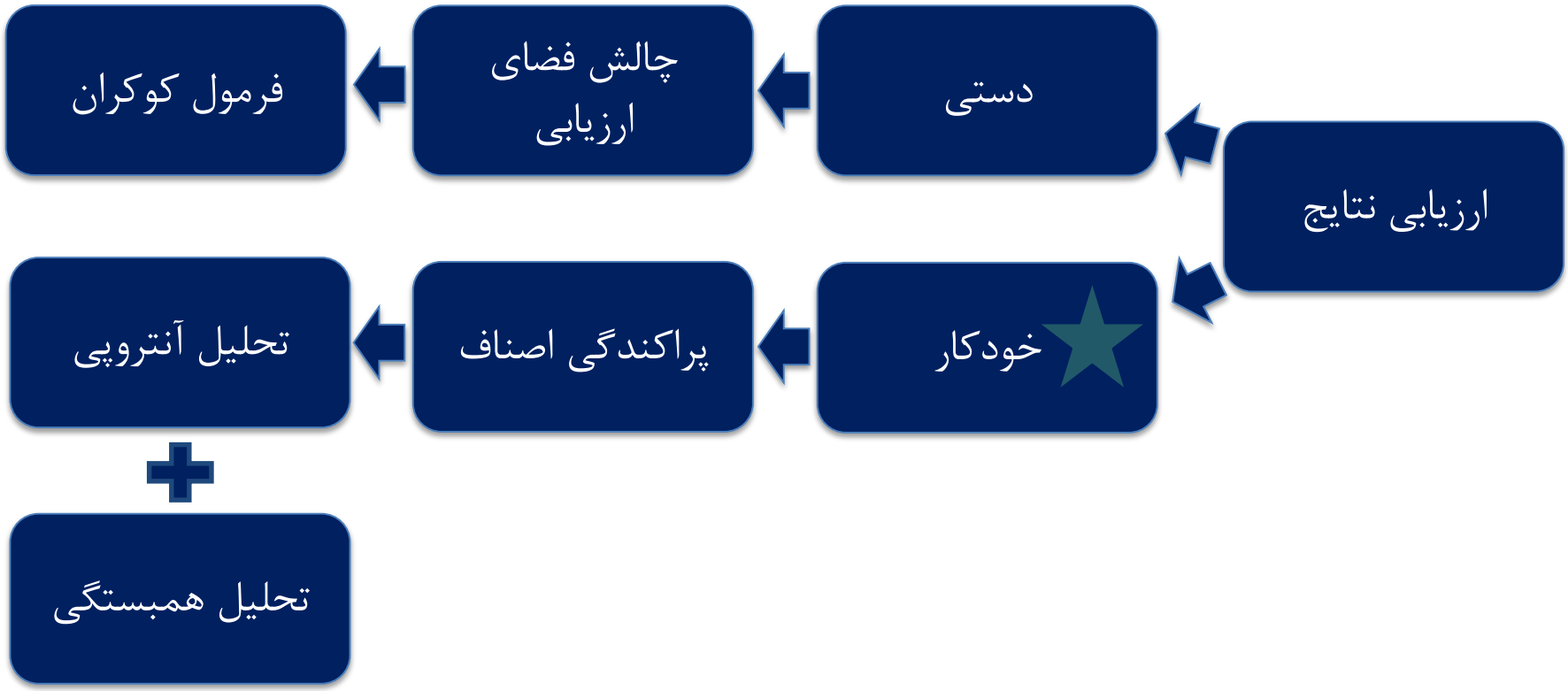
- جامعه آماری نتایج نهایی تشخیص و پیش‌بینی به دست آمده بزرگ است و نمی‌توان همه‌ی پذیرنده‌ها را بررسی دستی کرد.
- به‌همین خاطر از فرمول کوکران برای بدست آوردن حجم نمونه استفاده می‌کنیم.



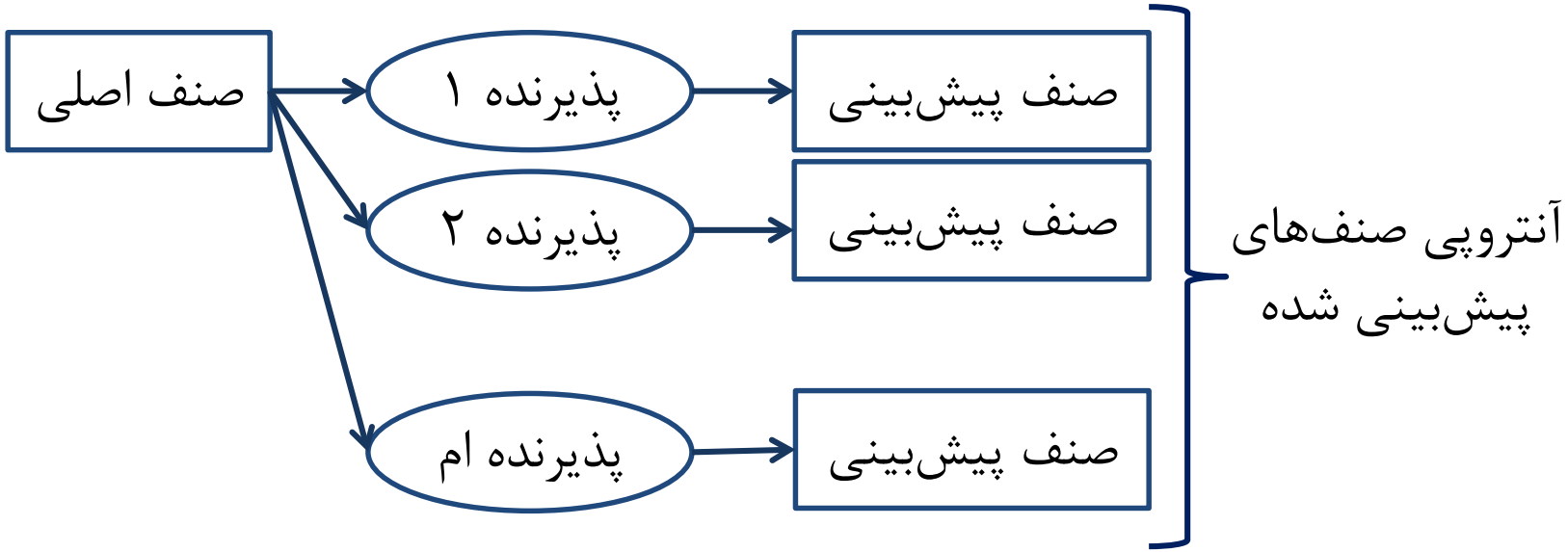
پیش‌بینی صنف جایگزین

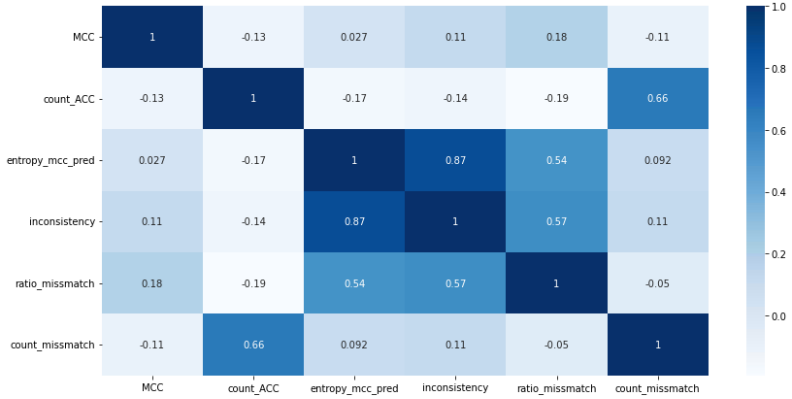
تشخیص ناهنجاری صنف

ارزیابی نتایج



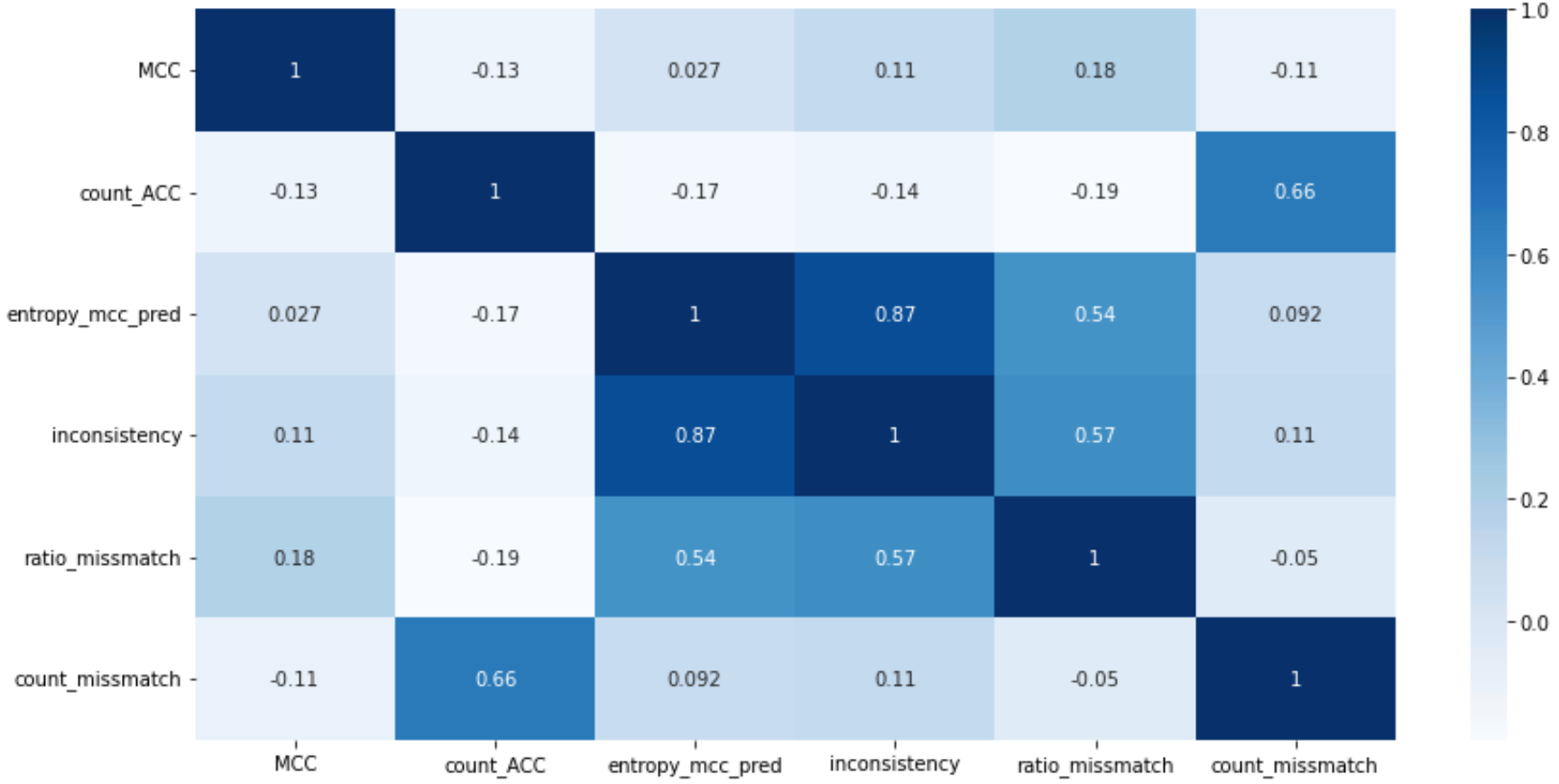
ارزیابی خودکار نتایج





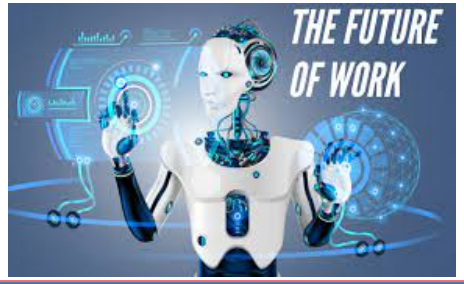
میزان همبستگی بین
آنتروپی صنف‌ها و عدم
تطابق پذیرندگان صنف‌ها

میزان همبستگی



نتیجه‌گیری

- نتیجه‌گیری
  - ارزش افزوده در تحلیل نام پذیرندگان
  - روش: نوآوری
    - ترکیب گراف‌کاوی و متن‌کاوی
  - ارزیابی: نوآوری
    - ترکیب آنتروپی و ضریب همبستگی



شرکت شبکه الکترونیکی پرداخت کارت (سهامی خاص)
شاپرک

با تشکر از وقت و توجه شما

بهناز پورابراهیم