

بسمه تعالی



دانشگاه صنعتی نوشیروانی بابل

عنوان طرح تحقیقاتی:

دسته‌بندی دانشجویان خوابگاه دانشگاه نوشیروانی بابل با کمک تکنیک خوشه‌بندی یادگیری ماشین

دانشجو:

علی مژده‌هی فرد

استاد راهنما:

دکتر مهدی حمیدی

تابستان ۱۴۰۱

## ۱- چکیده:

دانشگاه مکان مهم برای تبادل و تعامل فرهنگ‌های مختلف می‌باشد. دانشجویان زیادی در مقطع‌های تحصیلی متفاوت و رشته‌های گوناگون از مناطق مختلف کشور جهت فراگیری علم دور هم جمع می‌شوند. بسیاری از دانشگاه‌ها خدماتی رفاهی به دانشجویان غیر بومی خود اختصاص می‌دهند که به عنوان مهمترین آنها میتوان به سراهای دانشجویی اشاره کرد. دانشجویان غیر بومی عمدتاً سال‌های زیادی از تحصیل خود را در این خوابگاه‌ها می‌گذرانند. این خوابگاه‌ها با برطرف کردن دغدغه محل زندگی فرصت و بستر مناسبی را برای آرامش دانشجویان فراهم می‌کند تا از سال‌های تحصیل خود به بهترین شکل استفاده نمایند. با این حال اختلافات فرهنگی، اجتماعی و مذهبی میتواند باعث مشکلاتی در داخل خوابگاه گردد. دانشجویان با تیپ‌های شخصیتی متفاوت و فرهنگ‌های متفاوت علاوه بر اینکه می‌توانند به رشد فردی و اجتماعی هم کمک کنند میتوانند باعث به وجود آمدن تنش و نا ملایمت‌هایی در این مکان‌ها شوند. هدف از این طرح تحقیقاتی توسعه الگوریتم دسته‌بندی<sup>۱</sup> با کمک یادگیری ماشین<sup>۲</sup> برای دسته بندی دانشجویان بر اساس پارامترهای فردی به منظور کاهش تنش‌ها و افزایش رفاه روانی در محیط خوابگاه‌ها و اتاق‌ها می‌باشد.

## ۲- تعریف مسئله:

یادگیری ماشین زیر مجموعه‌ای از هوش مصنوعی می‌باشد و هدف آن استفاده از مدل‌ها و الگوریتم‌های آماری به منظور اتخاذ تصمیم توسط ماشین می‌باشد [۱]. امروزه از یادگیری ماشین در زمینه‌های زیادی از جمله مهندسی، پزشکی حتی سرگرمی استفاده می‌شود. موتور جستجو و سیستم رتبه دهی گوگل<sup>۳</sup>، یاهو<sup>۴</sup> سیستم پیشنهاد دهی شبکه‌های اجتماعی مانند فیسبوک<sup>۵</sup>، لینکدین<sup>۶</sup> و یا تکنولوژی تشخیص چهره همه از کاربردی یادگیری ماشین می‌باشند. روش‌های یادگیری ماشین را میتوان به دو دسته کلی بانظارت<sup>۷</sup> و بدون نظارت<sup>۸</sup> تقسیم بندی میکنند. در روش‌های با نظارت

---

<sup>۱</sup> Clustering

<sup>۲</sup> Machine learning

<sup>۳</sup> Google

<sup>۴</sup> Yahoo

<sup>۵</sup> Facebook

<sup>۶</sup> LinkedIn

<sup>۷</sup> Supervised

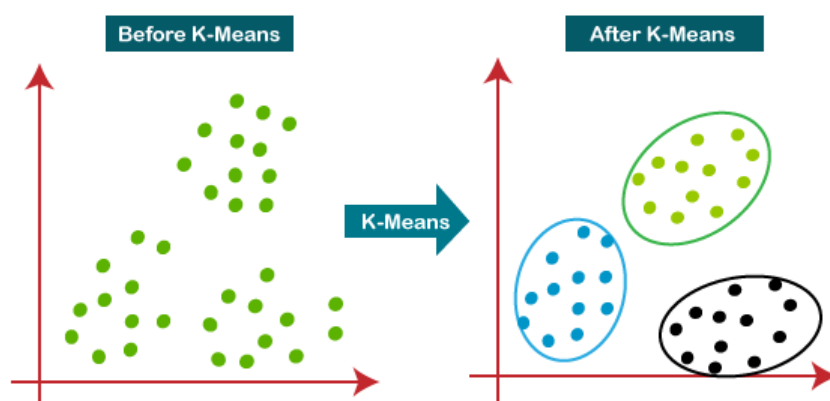
<sup>۸</sup> unsupervised

الگوریتم با یک سری داده برچسب گذاری شده آزمایشی آموزش میبندد و سپس به تحلیل سایر داده‌ها میپردازد مسئله‌های طبقه‌بندی<sup>۹</sup> و رگرسیون<sup>۱۰</sup> از این دسته می‌باشند. در روش بدون نظارت مانند خوشه بندی<sup>۱۱</sup> الگوریتم‌ها برای تصمیم گیری نیاز به برچسب گذاری از قبل ندارند. در این پروژه تحقیقاتی از روش‌های دسته بندی یادگیری ماشین استفاده میشود. از این روش‌ها میتوان به روش k-means روش سلسله مراتبی<sup>۱۲</sup> و روش BDSCAN نامبرد.

در این پژوهش سعی شده الگوریتم دسته بندی بر مبنای یادگیری ماشین توسعه داده شود. هدف این الگوریتم دریافت مشخصات فردی دانشجویان شامل نزدیک ترین محل سکونت، تیپ شخصیتی حاصل از تست mbti و عادت‌های شخصی شامل بهترین زمان مطالعه هر فرد (روز یا شب) و عادت سیگار کشیدن و یا نکشیدن دسته بندی کرد. در این پژوهش تمرکز اصلی بر توسعه الگوریتم می‌باشد و پارامترهای فردی و شخصی مورد استفاده هر دانشجو کاملاً برای مثال بوده و قابل تغییر و بهبود میباشد.

### ۳- مبانی نظری تحقیق:

در مسائل دسته بندی یادگیری ماشین، هدف دسته بندی داده‌ها در گروه‌هایی می‌باشند به شکلی که داده‌های در هر گروه بیشترین مشابهت را داشته باشند و همزمان گروه‌های مختلف بیشترین تفاوت را داشته باشند. یکی از روش‌های مورد استفاده برای دسته بندی استفاده از روش k-means می‌باشد. در این روش مشابهت داده‌ها را با رابطه فاصله اقلیدسی میسنجند [۲].



شکل ۱ دسته بندی داده‌ها با استفاده از روش k-means

<sup>۹</sup> classification  
<sup>۱۰</sup> Regression  
<sup>۱۱</sup> Clustering  
<sup>۱۲</sup> hierarchical

رابطه مورد استفاده برای محاسبه مشابهت در روش k-means برای دو داده فرضی شامل پارامترهای مختلف به شکل زیر می باشد:

$$Data^1: x_1, x_2, x_3, \dots$$

$$Data^2: x_1, x_2, x_3, \dots$$

$$Dis \left( Data^1, Data^2 \right) = \sqrt{\sum_{i=1}^n \left( x_{i1} - x_{i2} \right)^2}$$

هدف در این روش دسته بندی کردن داده ها به شکلی است که بهینه ترین حالت دسته بندی بر اساس فاصله اقلیدسی صورت بگیرد. به شکلی که تفاوت دسته ها ماکزیمم و تفاوت داده های هر گروه کمینه شود. در پژوهش حاضر که برای بلوک ۳ خوابگاه پسرانه دانشگاه نوشیروانی بابل در نظر گرفته شده تعداد ۱۹۶ دانشجو با داده های اتفاقی در نظر گرفته می شوند. هدف این است که ابتدا این ۱۹۶ دانشجو بر اساس مشابهت شخصیت و شهر محل زندگی در سه طبقه و شامل ۷ ردیف تقسیم شوند سپس بر اساس عادات فردی مثل ساعات مطالعه و سیگار کشیدن به صورت ۴ نفری در اتاق های هر ردیف (۷ اتاق) تقسیم می شوند. به عنوان نمونه اطلاعات هر دانشجو به صورت زیر نمایش داده می شود:

سیگار	شب کار یا روز کار	تایپ شخصیتی	شهر	کد دانشجو
نمیکشد	شب	ANFP-A	مشهد	۱
میکشد	روز	ENFP-T	رشت	۲

که در فایل ورودی برنامه اطلاعات برای هر دانشجو به صورت اعدادی به شکل زیر نمایش داده می شوند:

سیگار	شب کار یا روز کار	تایپ شخصیتی	شهر	کد دانشجو
۰	۱	۹	۱۱	۱
۱	۰	۱۰	۲۵	۲

برای ثبت اطلاعات هر دانشجو یک وبسایت توسعه داده می شود به این شکل که اطلاعات مورد نیاز هر دانشجو را دریافت کرده و در نهایت داده ها را به شکل یک فایل CSV برای شروع برنامه تقسیم بندی دانشجویان آماده می کند.

#### ۴- فرضیات تحقیق:

- ۱- مدل نوشته شده با توجه به مشخصات بلوک ۳ خوابگاه پسرانه دانشگاه نوشیروانی بابل در نظر گرفته شده است.
  - ۲- جنسیت همه دانشجویان یکسان در نظر گرفته شده است.
  - ۳- تحصیلات همه در مرحله تکمیلی می باشند. (ارشد و دکتری)
  - ۴- در قسمت هر شهر نزدیک ترین مرکز استان به شهر محل سکونت در نظر گرفته شده است.
- کد هر شهر با توجه به جدول زیر مشخص می شود:

مرکز استان	استان	کد
تبریز	آذربایجان شرقی	۱
ارومیه	آذربایجان غربی	۲
اردبیل	اردبیل	۳
اصفهان	اصفهان	۴
کرج	البرز	۵
ایلام	ایلام	۶
بندر بوشهر	بوشهر	۷
تهران	تهران	۸
شهرکرد	چهارمحال و بختیاری	۹
بیرجند	خراسان جنوبی	۱۰
مشهد	خراسان رضوی	۱۱
بجنورد	خراسان شمالی	۱۲
اهواز	خوزستان	۱۳
زنجان	زنجان	۱۴
سمنان	سمنان	۱۵
زاهدان	سیستان و بلوچستان	۱۶
شیراز	فارس	۱۷
قزوین	قزوین	۱۸
قم	قم	۱۹
سنندج	کردستان	۲۰
کرمان	کرمان	۲۱
کرمانشاه	کرمانشاه	۲۲
یاسوج	کهگیلویه و بویراحمد	۲۳
گرگان	گلستان	۲۴

رشت	گیلان	۲۵
خرم‌آباد	لرستان	۲۶
ساری	مازندران	۲۷
اراک	مرکزی	۲۸
بندرعباس	هرمزگان	۲۹
همدان	همدان	۳۰
یزد	یزد	۳۱

۵- برای لحاظ کردن تیپ شخصیتی دانشجویان از تست روانشناسی mbti استفاده شده است که هر تیپ شخصیتی با کد مخصوص خود به شرح جدول زیر مشخص می‌شود:

کد تیپ شخصیتی	تیپ شخصیتی
۱	تیپ شخصیتی ENFJ: برون‌گرا - شهودی - احساسی - قضاوتی
۲	تیپ شخصیتی ENFP: برون‌گرا - شهودی - احساسی - ادراکی
۳	تیپ شخصیتی ESFJ: برون‌گرا - حسی - احساسی - قضاوتی
۴	تیپ شخصیتی ESFP: برون‌گرا - حسی - احساسی - ادراکی
۵	تیپ شخصیتی INFJ: درون‌گرا - شهودی - احساسی - قضاوتی
۶	تیپ شخصیتی INFP: درون‌گرا - شهودی - احساسی - ادراکی
۷	تیپ شخصیتی INTJ: درون‌گرا - شهودی - منطقی - قضاوتی
۸	تیپ شخصیتی INTP: درون‌گرا - شهودی - منطقی - ادراکی
۹	تیپ شخصیتی ISFJ: درون‌گرا - حسی - احساسی - قضاوتی
۱۰	تیپ شخصیتی ISFP: درون‌گرا - حسی - احساسی - ادراکی
۱۱	تیپ شخصیتی ISTJ: درون‌گرا - حسی - منطقی - قضاوتی
۱۲	تیپ شخصیتی ISTP: درون‌گرا - حسی - منطقی - ادراکی
۱۳	تیپ شخصیتی ESTJ: برون‌گرا - حسی - منطقی - قضاوتی
۱۴	تیپ شخصیتی ENTJ: برون‌گرا - شهودی - منطقی - قضاوتی
۱۵	تیپ شخصیتی ENTP: برون‌گرا - شهودی - منطقی - ادراکی
۱۶	تیپ شخصیتی ESTP: برون‌گرا - حسی - منطقی - ادراکی

۶- برای عادات فردی اگر فرد مایل به کار و مطالعه در روز باشد عدد صفر و اگر مایل به کار در شب باشد عدد یک، همچنین اگر مایل به استعمال سیگار باشد عدد یک، و در غیر اینصورت عدد صفر برای این پارامترها لحاظ می‌شوند.

## ۵- برنامه زمانبندی:

برای آماده سازی برنامه مورد نظر گام‌های زیر مد نظر می‌باشد.

۴ روز	۱. مطالعات پیش نیاز جهت توسعه وبسایت و الگوریتم‌های مورد استفاده
یک هفته	۲. توسعه وبسایت جهت دریافت اطلاعات دانشجویان
یک هفته	۳. توسعه الگوریتم دسته بندی بر اساس شهر و تیپ شخصیتی
یک هفته	۴. توسعه الگوریتم تقسیم بندی دانشجویان در اتاق بر اساس خصیصه‌های فردی و عیب یابی الگوریتم کلی

## ۶- هزینه:

با توجه به گام‌های مطرح شده برای پروژه هزینه به دو قسمت و به شکل زیر تعیین می‌شود:

۱. هزینه طراحی و اجرای وبسایت (۴ میلیون تومان)
۲. هزینه توسعه الگوریتم دسته بندی و عیب یابی کلی سیستم (۴ میلیون تومان)

## ۷- منابع:

- 1- Hastie, Trevor; Tibshirani, Robert (2009). *The Elements of Statistical Learning: Data mining, Inference, and Prediction*. New York: Springer. pp. 485–586.
- 2- Estivill-Castro, Vladimir (20 June 2002). "Why so many clustering algorithms – A Position Paper". *ACM SIGKDD Explorations Newsletter*. **4** (1): 65–75.