بسمه تعالی

**فرم ثبت نام در چالش**

**"داده کاوی هوشمند به توان شناختی"**

**سال 1398**

# معرفی تیم شرکت کننده در چالش

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ردیف** | **نام و نام خانوادگی** | **نقش در تیم** | **تخصص در تیم** | **رشته تحصیلی** | **آخرین مدرک تحصیلی** | **دانشگاه** | **موبایل** | **ایمیل** |
| **۱** | **محمدامین خدامرادی** | **مدیر تیم** | **داده کاوی و برنامه‌نویسی** | **علوم کامپیوتر** | **کارشناسی ارشد** | **شهیدبهشتی** | **09361643465** | [**Khodamoradi1992@gmail.com**](mailto:Khodamoradi1992@gmail.com) |
| **۲** | **سمیه گل‌محمدی** | **رابط تیم** | **اقتصادسنجی و برنامه‌نویسی** | **علوم کامپیوتر** | **کارشناسی ارشد** | **علامه طباطبایی** | **09358309122** | [**So\_gmo@yahoo.com**](mailto:So_gmo@yahoo.com) |

**اطلاعات کسب و کار نوپا یا شرکت دانش بنیان در صورت عضویت تیم:**

|  |  |
| --- | --- |
| **نام کسب و کار** |  |
| **آدرس وب سایت کسب و کار** |  |
| **سال تشکیل کسب و کار** |  |
| **نوآوری/محصول /خدمتی که کسب و کار ارائه می‌دهد** |  |

# توصیف اجمالی طرح

|  |
| --- |
| **الف-مرور پژوهش (حداکثر 2000 کلمه):** |
| **ب-تشریح ایده (حداکثر 2000 کلمه):**  **در اين پژوهش برآنیم از مجموعه داده‌ای که توسط شرکت Tap30 برای این چالش فراهم شده استفاده نماییم. بعد از مطالعه و بررسی داده‌ها به این نتیجه رسیدم که می‌توان تحلیل‌هایی بصری از رفتار راننده‌ها و مسافران برای این مجموعه داده ارائه داد. با به کارگیری ابزارهای مصورسازی و زبا‌ن‌های برنامه‌نویسی هم‌چون R و Pythoy در صدد خواهیم بود این تحلیل‌ها را فراهم و ارائه دهیم. تحلیل‌ها به دو دسته‌ی مبتنی بر زمان و بدون درنظر گرفتن زمان تقسیم می‌شود.**  **این تحلیل‌ها در جهت پاسخگویی به سوالاتی که در ادامه ذکر خواهد شد ارائه می‌شوند:**   1. **نحوه‌ی توزیع مکان اولیه‌ی مسافران برروی نقشه چگونه است؟** 2. **نحوه‌ی توزیع مقاصد مسافران برروی نقشه چگونه است؟** 3. **توزیع انواع سفرهای کنسل شده، راننده یافت نشد و به اتمام رسیده روی نقشه چگونه است؟ آیا رابطه‌ی معناداری بین موقعیت جغرافیایی و وضعیت نهایی سفر وجود دارد؟** 4. **نمودار حراراتی درخواست سفر در روزهای هفته و ساعت‌های روز به چه شکل بوده؟ و زمان‌های کم درخواست و پر درخواست در بازه‌ی یک هفته چه زمان‌هایی هستند؟** 5. **اگر داده‌ی مربوط به وضعیت آب و هوا را داشته باشیم، رفتار تقاضای مسافران و نحوه‌ی درخواست دادن آن‌ها در وضعیت‌های مختلف آب‌و‌هوایی چگونه خواهد بود؟ و چه مقدار تغییر خواهد کرد؟** 6. **رفتار مسافرانی که سفر آن‌ها توسط راننده قبول شده چگونه است؟ چه مسافرانی، سفر خود را لغو می‌کنند؟ یک مدل یادگیری ماشین می‌تواند این رفتار را تحلیل کرده و پیش بینی نماید.** 7. **اگر داده‌های سفرهای پیشنهادی به هر راننده را داشته باشیم، آیا می‌توان پیش بینی کرد کدام راننده یک درخواست سفر را قبول می‌کند؟ یک مدل یادگیری ماشین می‌تواند این رفتار را تحلیل و پیش بینی نماید.** 8. **اگر اطلاعات زمان پایان آخرین سفر قبلی راننده را داشته باشیم، آیا زمان انتظار راننده به عنوان یک المان و ویژگی می‌تواند در مدل یادگیری ماشین درنظر گرفته‌شده و مدل را تقویت کند؟**   **پاسخ‌گویی به سوالات مذکور و ارائه نمودارها‌ و شکل‌های خواسته‌شده می‌تواند نقش مهمی در تحلیل رفتاری مسافران و رانندگان داشته‌باشد. به گونه‌ای که برای هر روز هفته و ساعت‌های روز با افزایش و یا کاهش موقتی قیمت‌ها باعث جذب و یا دفع راننده‌ها شود. یا با استفاده از پیام‌های نرم‌افزار، راننده‌هایی که زمان زیادی منتظر بوده‌اند را از مکان‌هایی که تقاضا بیش از عرضه است مطلع کند.**  **همچنین می‌توان سیستم پیشنهاد سفر به راننده را تغییر داده و با درنظر گرفتن المان‌های بیشتر نسبت به قبل تقویت نمود.**  **داده‌های زیر همان‌طور که در متن شرح داده شد، درصورت وجود و دراختیار بودن می‌توانند تحلیل‌ها را تقویت کرده و در ارائه‌ی مدل رفتاری بهتر از مسافران و رانندگان کمک کنند:**   1. **داده‌ی مربوط به وضعیت آب و هوا در زمان‌ درخواست سفر** 2. **داده‌ها‌ی سفرهای پیشنهادی به هر راننده** 3. **اطلاعات زمان پایان آخرین سفر قبلی راننده (زمان انتظار راننده برای سفر بعدی)** |
| **ج-تشریح کیفیت و مراحل اجرای طرح (حداکثر 1000 کلمه):**  **ابتدا لازم است، داده‌ها به طور کامل از شرکت Tap30 دریافت شده و درمورد سیستم محاسبه‌ی driverETA اطلاعات کافی جمع‌آوری شود. سپس داد‌هایی که مقدار آن‌های گذاشته شنده یا مقادیر غیرمتعارف دارند سازماندهی و درصورت لزوم حذف شوند. داده‌های به شکلی آماده شوند که بتوان نمودار‌ها و شکل‌های مذکور را از روی آن‌ها رسم کرده و ارائه داد. سپس تحلیل‌های هر شکل را بیان کرده و این تحلیل‌های براساس روابط آماری اعتبارسنجی شده و معنی‌دار می‌شوند. در مرحله‌ی بعدی مدل‌های یادگیری ماشین مختلف اعمال شده و مدلی با دقت و توان پیش‌بینی بالاتر برروی داده‌های آزمون معرفی می‌شود و با بقیه‌ی مدل‌ها مقایسه خواهد شد. جهت سنجش مقدار قابل اعتماد بودن مدل و ثبات آن ، مدل در یک ساز و کار اعتبار سنجی متقابل (kfold\_CV) بر روی داده‌های آزمون، آزموده می‌شود. معیارهایی هم‌چون AUC، AUPR و Fmesear ملاک هاي ارزيابي مدل يادگيري ماشين خواهند بود.** |
| **د-دادگان به کار گرفته شده در این طرح :**  **مجموعه داده‌ی Tap30** |

# تعهدات

اینجانب محمدامین خدامرادی به عنوان مدیر تیم با امضای این فرم کلیه مندرجات آن­را از نظر صحت و دقت تایید می­نمایم.

|  |
| --- |
| نام و نام خانوادگی: محمدامین خدامرادی امضا: تاریخ: ۲۴/۰۸/۱۳۹۸ |