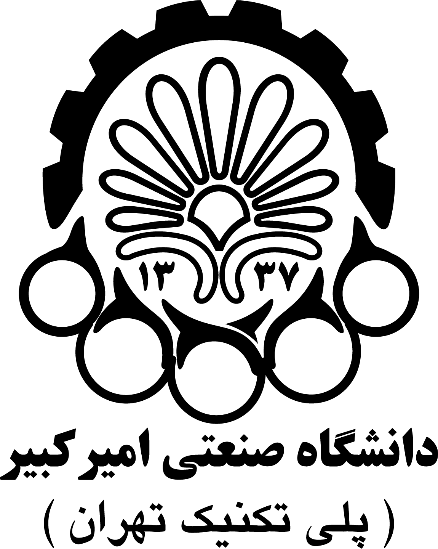
****

**دانشکده مهندسي کامپیوتر و فناوری اطلاعات**

گزارش پروژه نهایی درس کاربرد تحلیل شبکه ای در تجارت الکترونیک:

**آنالیز شبکه ای روی دادگان سفرهای هوایی ناوگان ایالات متحده**

**ارائه دهنده:**

**محمد چرم زاده**

**استاد راهنما:**

**دکتر هاشمی گلپایگانی**

**تابستان 98**

فهرست مطالب

**عنوان صفحه**

[معرفی دادگان 1](#_Toc106512875)

[پرسش های طرح شده 3](#_Toc106512877)

[سناریوهای متناظر 3](#_Toc106512877)

[روش آماده سازی دادگان 3](#_Toc106512877)

[سناریو 1 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

[سناریو 2 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

[سناریو 3 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

[سناریو 4 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

[سناریو 5 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

[سناریو 6 6](#_Toc106512881)

[استخراج شبکه 7](#_Toc106512882)

[3-روش تحلیل 7](#_Toc106512883)

[3-2-1. نتایج 7](#_Toc106512884)

معرفی دادگان

دادگانی که در این پروژه روی آن متمرکز هستیم مربوط به اطلاعات پروازهای انجام شده در ناوگان حمل و نقل هوایی ایالات متحده است که شامل 4 جدول به شرح :

**airlines.csv**  **: اطلاعات شرکت های ایرلاین**

**airports.csv**  **: اطلاعات فرودگاه های ایالات متحده**

**flights.csv** **: پروازهای انجام شده در محدوده منتهی به سال 2015**

**Airports2.csv**  **: اطلاعات پروازهای انجام شده با تمرکز بر اطلاعات مسافرین**

**همچنین هر یک از این جداول دارای ستون هایی به شرح جدول زیر می باشد:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **نام فایل/جدول** | **عنوان فیلد** | **نوع فیلد** | **شرح محتوای فیلد** |
| airlines.csv | IATA\_CODE | string | Airline Identifier |
| airlines.csv | AIRLINE | string | Airport's Name |
| airports.csv | IATA\_CODE | string | Location Identifier |
| airports.csv | AIRPORT | string | Airport's Name |
| airports.csv | CITY | string | Airport's City |
| airports.csv | STATE | string | Airport's State |
| airports.csv | COUNTRY | string | Airport's country |
| airports.csv | LATITUDE | numeric | Latitude of the Airport |
| airports.csv | LONGITUDE | numeric | Longitude of the Airport |
| flights.csv | YEAR,MONTH,DAY,DAY\_OF\_WEEK | Date | flight's Date |
| flights.csv | FLIGHT\_NUMBER | string | Flight Identifier |
| flights.csv | TAIL\_NUMBER | string | Aircraft Identifier |
| flights.csv | ORIGIN\_AIRPORT | string | Starting Airport |
| flights.csv | DESTINATION\_AIRPORT | string | Destination Airport |
| flights.csv | SCHEDULED\_DEPARTURE | numeric | Planned Departure Time |
| flights.csv | DEPARTURE\_TIME | numeric | WHEEL\_OFF - TAXI\_OUT |
| flights.csv | DEPARTURE\_DELAY | numeric | Total Delay on Departure |
| flights.csv | DISTANCE | numeric | Distance between two airports |
| flights.csv | CANCELLED | boolean | Flight Cancelled (1 = cancelled) |
| Airports2.csv | Passengers | numeric | Number of passengers transported from origin to destination |
| Airports2.csv | Seats | numeric | Number of seats available on flights from origin to destination |
| Airports2.csv | Origin\_population | numeric | Origin city's population as reported by US Census |
| Airports2.csv | Destination\_population | numeric | Destination city's population as reported by US Census |

**پرسش های مورد نظر**



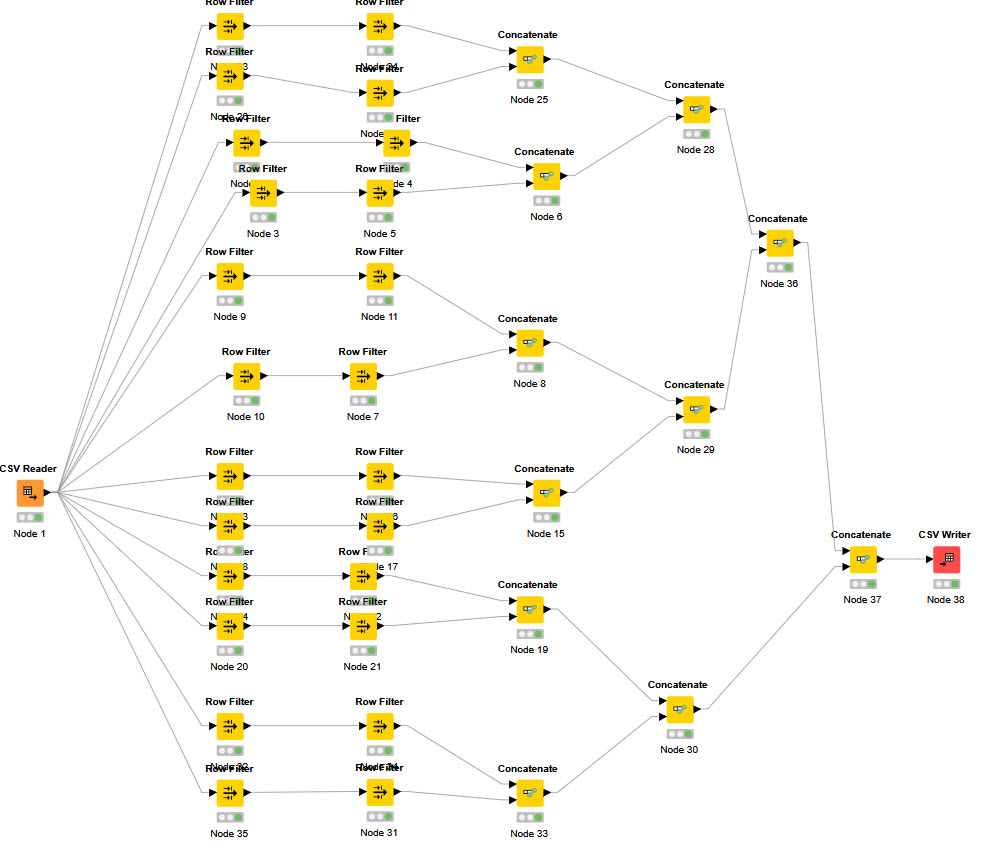
سناریوهای متناظر

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **id** | **scenario titile** | **nodes** | **edges** | **weight** | | 1 | تاخیر ایرلاین ها در سال 2015 | ایرلاین ها | داشتن پرواز مشترک  در یک خط هوایی | تفاضل مقدار تاخیر  دو ایرلاین در همان مسیر | | 2 | ظرفیت خالی به پر شده | فرودگاه ها | انجام سفر پروازی | مقدار ظرفیت پر نشده - کل ظرفیت | | 3 | وضعیت جهان کوچک بودن سفرها | فرودگاه ها | انجام سفر پروازی | تعداد سفرها | | 4 | فرودگاه های هاب | فرودگاه ها | انجام سفر پروازی | تعداد سفرها | | 5 | تاخیر مسیرهای پروازی | فرودگاه ها | انجام سفر پروازی | مجموع تاخیر در پروازهای هر مسیر | | 6 | فرودگاه های پر پرواز (جهت دار) | فرودگاه ها | فرودگاه ها | تعداد سفرها | |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

آماده سازی دادگان

**دادگان اولیه دارای رکورد های بیش از 4 میلیون و حجمی بیش از 1GB بودند که برای انجام پردازش ها و تحلیل های معمول ناچار به کاهش هدفمند و سپس** cleansing **آن ها بودیم.**

**برای این منظور ابتدا دادگان پر حجم پروازها را با استفاده از KNIME به 12 ماه سال تفکیک کرده و سپس از هر ماه به صورت رندوم و تصادفی به انتخاب 1000 رکورد پرداختیم. سپس ردیف هایی که ستون های اصلی و مورد نظر ما در آن ها ناقص بود را حذف کردیم و به جای آن ها ردیف های مشابه را جایگزین کردیم.**

****

سناریو 1

استخراج شبکه

گره ها در این شبکه همان ایرلاین ها می باشند که از فایل **airlines.csv و پس از حذف اطلاعات در node-s1.csv قرار گرفتند. جهت افزایش سرعت و اتوماتیک کردن انتقال اطلاعات برنامه پایتونی node-S1 نوشته و فایل گره ها توسط آن تولید شده است**

**برای حصول یال ها با شرط داشتن پرواز در یک مسیر مشترک اسکریپت پایتونی edge-S1 نوشته شد و در هر بار حصول یک یال ، وزن آن نیز که عبارت است از تفاضل قدرمطلقی تاخیر نیز محاسبه و در ستون وزن وارد شد.**

**روش تحلیل**

**با توجه تعریف اولیه سناریو ، پارامتر مورد نظر ما در شبکه حاصل همان وزن یال های بین نود ها هستند یعنی در واقع بیشترین نودی که دارای یال های با بیشترین وزن است دارای تاخیر بیشتری است و اینگونه ایرلاین های با بیشترین تاخیر پروازی شناسایی می شوند.**

**نتایج**

**با توجه به شبکه حاصل و محاسبه پارامتر های وزن نود ها، ایرلاین های با بیشترین تاخیر در بازه زمانی یکسان بدست امدند.**

سناریو 2

استخراج شبکه

گره ها در این شبکه همان فرودگاه ها می باشند که از فایل **airports.csv و پس از حذف اطلاعات در node-s2.csv قرار گرفتند. جهت افزایش سرعت و اتوماتیک کردن انتقال اطلاعات برنامه پایتونی node-S2 نوشته و فایل گره ها توسط آن تولید شده است**

**برای حصول یال ها با شرط داشتن پرواز در یک مسیر مشترک اسکریپت پایتونی edge-S2 نوشته شد و در هر بار حصول یک یال وزن آن که شامل** مقدار ظرفیت پر نشده - کل ظرفیت **بوده محاسبه و در ستون وزن وارد می شود.**

**روش تحلیل**

**در این شبکه وزن یال ها که نشانگر مقدار ظرفیت پر نشده - کل ظرفیت است می تواند شاخصی برای یافتن**