



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

پایان نامه ی کارشناسی

عنوان:

دسته بندی خانوارهای ایرانی براساس وضعیت
رفاهی - اقتصادی و ارائه معیاری برای سنجش کیفیت
دهک بندی

نگارش:

محمدصدرا حیدری - علیرضا ایلامی

استاد راهنما:

جناب آقای دکتر محمدامین فضلی

شهریور ۱۴۰۱

فصل ۱

دادگان

۱-۱ معرفی داده

در این پروژه از داده‌های عمومی «پایگاه اطلاعات رفاه ایرانیان» استفاده می‌کنیم. این پایگاه اطلاعات نتیجه تلاش وزارت تعاون، کار و رفاه اجتماعی جمهوری اسلامی ایران برای ایجاد شناسنامه رفاهی-اقتصادی افراد است.

پایگاه اطلاعات رفاه ایرانیان، از تجمیع ۵۰ منبع داده‌ای تشکیل شده است. در فاز نخست، ۲۵ منبع داده‌ای به طور کامل تجمیع و ساختار اصلی پایگاه اطلاعات رفاه ایرانیان (با بیش از ۶۰ جدول داده‌ای) بر اساس آن‌ها شکل گرفته است. بیش از ۳ میلیارد رکورد داده‌ای در پایگاه داده نهایی ذخیره و تجمیع شده است. در حال حاضر، ۲۲۱ فیلد اطلاعاتی شناسایی شده‌اند که ابعاد مختلف هویتی شهروندان را به صورت مستقیم و غیرمستقیم توصیف می‌نمایند. بخشی از خروجی‌های پایگاه اطلاعات رفاه ایرانیان به صورت وب‌سرویس در اختیار ۲۰ سازمان مختلف قرار گرفته است. علاوه بر آن، خدمات تحلیلی نیز به تعدادی از سازمان‌ها ارائه شده است. این خدمات شامل استحقاق سنجی، بررسی همپوشانی، داده‌کاوی، وسع سنجی و ... می‌شود. (لینک به صفحه)

پایگاه ملی اطلاعات رفاه ایرانیان حاوی اطلاعات مفیدی در مورد ابعاد مختلف اجتماعی-اقتصادی شهروندان ایرانی است که می‌تواند برای اهداف مختلف مورد تحلیل و بررسی اندیشمندان، جامعه‌شناسان، اقتصاددانان و به طور کلی جامعه محققان کشور قرار گیرد. از سمتی دیگر، این پایگاه شامل اطلاعات

محرمانه و خصوصی شهروندان است که نقض این حریم خصوصی، ممکن است تبعات جبران ناپذیری در زمینه اعتماد عمومی ایجاد نماید. معاونت رفاه اجتماعی به منظور ایجاد یک مصالحه میان ۱ - امکان دسترسی جامعه محققان به داده‌های پایگاه اطلاعات ایرانیان جهت انجام مطالعات تحقیقاتی و ۲ - حفظ حداکثری حریم خصوصی شهروندان ایرانی، اقدام به تهیه نمونه‌های ۲ درصدی از اطلاعات موجود در پایگاه ملی اطلاعات رفاه ایرانیان نموده است. در این نمونه‌برداری‌ها سعی شده است که نکات زیر به صورت جدی مدنظر قرار گیرند.

۱-۲ معرفی ویژگی‌ها

دیتاست موجود شامل اطلاعات پانصدهزار خانوار و ۱/۴۹۰/۹۹۱ فرد ایرانی است و برای هر فرد ۴۸ ویژگی دارد. ویژگی‌های موجود در دیتاست را می‌توان به دسته‌های زیر تقسیم‌بندی کرد.

• اطلاعات شخصی

- شناسه فرد
- تاریخ تولد
- شناسه سرپرست خانوار
- جنسیت

• محل زندگی

- کد پستی
- شهرستان محل زندگی
- استان محل زندگی
- شهری یا روستایی بودن

• اطلاعات حساب (برای سال‌های ۹۵ تا ۹۸ و به تفکیک سال)

- گردش بستانکار
- مانده ابتدای سال
- گردش یده‌کار
- مانده انتهای سال
- مجموع سود حساب‌ها

• اطلاعات تراکنش (سال ۹۸ و شش ماهه اول سال ۹۹ و به تفکیک سال)

- مقدار کل تراکنش کارت‌ها
- تعداد کل تراکنش کارت‌ها

• دارایی‌ها

- تعداد خودروهای فرد
- مجموع ارزش خودروهای فرد

• تفریحات (در بازه سال‌های ۹۶ تا ۹۹)

- تعداد سفرهای خارجی هوایی غیرزیارتي
- تعداد سفرهای خارجی زمینی غیرزیارتي
- تعداد سفرهای خارجی هوایی زیارتي
- تعداد سفرهای خارجی زمینی زیارتي

• کسب و کار

- داشتن مجوز صنفی
- صنفی که در آن مجوز دارد

• بیمه

- داشتن بیمه سلامت
- نوع بیمه سلامت

• شرایط خاص

- داشتن بیماری خاص
- معلول بودن فرد

• بازنشستگی

- بیمه‌پرداز صندوق‌های بازنشستگی بودن
- بازنشسته صندوق‌های بازنشستگی بودن

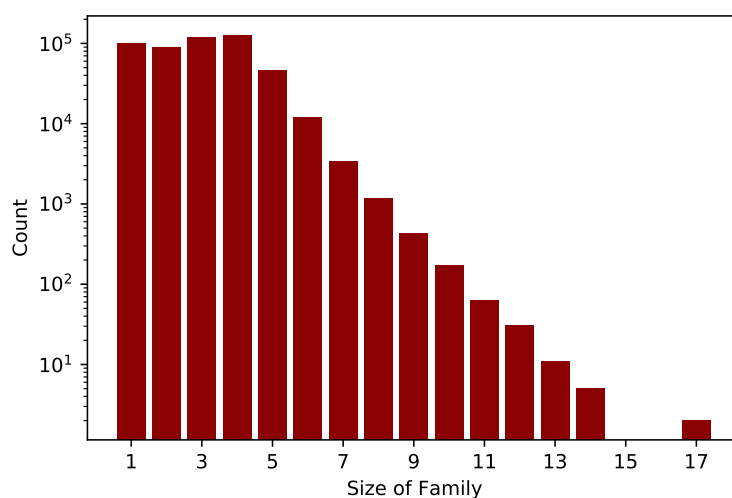
• مالیات و درآمد

- شاغل مشمول مالیات بودن
- مجموع درآمد فرد از حقوق

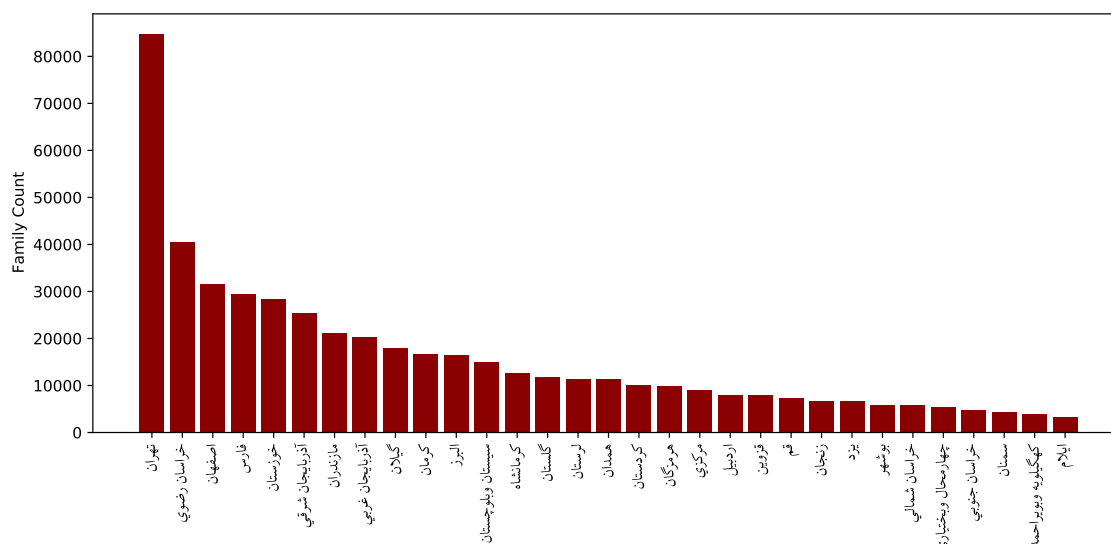
۱-۳ آمارها و شهود

۱-۳-۱ اندازه خانوار

۱-۳-۲ توزیع خانوارها در کشور



شکل ۱-۱: نمودار هیستوگرام اندازه خانوارهای موجود در نمونه ۲ درصدی پایگاه اطلاعات رفاه ایرانیان. در نمودار بالا محور عمودی به صورت لگاریتمی مقیاس شده است.



شکل ۱-۲: نمودار تعداد خانوارهای موجود در نمونه ۲ درصدی پایگاه اطلاعات رفاه ایرانیان به تفکیک استان محل زندگی. بیشترین تعداد خانوار مربوط به استان تهران و کمترین آن مربوط به استان ایلام می باشد.



Sharif University of Technology
Department of Computer Engineering

M.Sc. Thesis

Iranian Household Clustering and developing new metric to test quality of deciles

By:

Mohammad Sadra Heydari - Alireza Ilami

Supervisor:

Dr. Mohammad Amin Fazli

September 2020