

پروژه پیشنهادی شامل چندین فاز بوده که باید طبق ترتیب زمانی تعیین شده انجام شوند. با وجود اینکه فازهای اولیه برای تمام تیم ها مشترک است اما باید به طور جداگانه در هر تیم انجام گیرد.

فاز (۱) استفاده از ابزار داده کاوی WEKA (مهلت تحویل تا ۹۷/۲/۲۶)

الف) آشنایی با استفاده از ابزار WEKA جهت فراخوانی چندین تکنیک متفاوت یادگیری ماشین. آخرین نسخه تثبیت شده را استفاده کنید. هم از رابط گرافیکی (Explorer) و هم از رابط خط فرمان (CLI) استفاده کنید. به لینک <https://www.cs.waikato.ac.nz/~ml/weka> جهت دسترسی به مستند WEKA مراجعه کنید.

ب) تکنیک های یادگیری معرفی شده در زیر را با تنظیمات پیش فرض با هدف تحلیل داده های هوا (موجود در فایل weather.arff ضمیمه شده به این دستورالعمل) بکار بگیرید. برای حالت های تست، ابتدا "use training set" را انتخاب کنید و سپس "percentage split" با تعیین مقدار پیش فرض ۶۶٪ انتخاب کنید. سپس در صد نرخ خطای مدل بکار رفته را گزارش دهید. در ادامه این مقدار را تغییر دهید و نتایج بدست آمده را تحلیل کنید.

- ZeroR (majority class)
- OneR
- Naïve Bayes Simple
- J48

ج) کدام یک از این طبقه بندها به شما اطمینان بیشتری برای استفاده می دهد؟ توضیح دهید.

د) در مورد میزان دقت بدست آمده برای حالت تستی که از تمام داده های آموزشی استفاده می شود و حالتی که درصدی از آنها جدا می شود چه تحلیلی می توانید ارائه دهید؟