

# هوش مصنوعی و سیستم‌های خبره

نیمسال دوم 1399-1400

پروژه پایانی

## پروژه شماره 1: تشخیص زبان متن

### هدف از پروژه

هدف از این پروژه پیاده سازی مدلی است که با گرفتن متن به عنوان ورودی، زبان آن را تشخیص دهد. هر پروژه یک دسته بندی 6 کلاسه است که 3 کلاس آن از دسته زبانهای با الفبای لاتین و 3 کلاس دیگر از زبان های با الفبای فارسی-عربی هستند. برای همه دانشجویان دسته بندی میان انگلیسی المانی فارسی و عربی مشترک است و هرکس می تواند یک زبان از سایر زبانهای لاتین مثل فرانسه، هلندی، سوئدی و ... و یک زبان از سایر زبان های فارسی مثل پشتو، ترکی، کردی، تاجیکی و ... انتخاب کند تا 6 کلاسش تکمیل گردد.

## پروژه شماره 2: تبدیل کلمات انگلیسی به شکل فارسی نوشت آنها

### هدف از پروژه

هدف از این پروژه پیاده سازی مدلی است که با گرفتن کلمه انگلیسی، معادل فارسی آن را تولید کند. مجموعه دادگان تست این پروژه حوزه عمومی و حوزه تخصصی کامپیوتر خواهد بود.

به پیوست برای هر پروژه یک فایل ارایه می‌شود:

- فایل test1.csv: هر رکورد داده در این فایل یک جمله یا متن به یمنی از زبان های گفته شده می‌باشد. شما باید برای هرکدام زبان آن را پیشبینی کرده و به صورت یک فایل CSV آن را در سامانه بارگذاری کنید. سپس دقت شما توسط سیستم سنجیده شده و در جدول امتیازات باتوجه به امتیاز خود رتبه ای می‌گیرید.
- فایل test2.csv: هر رکورد داده در این فایل کلمه ای انگلیسی است که شما باید برای هرکدام معادل فارسی اش را تولید کنید. سپس در قالب یک فایل CSV آن را در سامانه بارگذاری کنید. سپس دقت شما توسط سیستم سنجیده شده و در جدول امتیازات باتوجه به امتیاز خود رتبه ای می‌گیرید.

در هر دو پروژه روش انجام پروژه آزاد است و دانشجویان می توانند با توجه به مطالب اموخته شده در کلاس روش کار را انتخاب کنند. در صورت نیاز به دادگان، تهیه آن برعهده دانشجویانست.

### سنجش دقت

در نهایت دقت کلاس بندی شما روی داده تست، (درصد داده هایی که وجود و یا عدم وجود توهین در آنها را درست تشخیص داده اید) توسط سیستم سنجیده شده می‌شود و با توجه به آن رتبه شما در scoreboard تعیین خواهد شد.

## فرمت فایل submission

فایلی که در لینک پروژه آپلود می‌کنید باید یک فایل csv با X سطر و یک سطر header باشد. در صورتی که تعداد ستون و یا سطرهای فایل ارسالی متفاوت باشد؛ با ارور مواجه خواهید شد. این فایل باید دقیقا 2 ستون داشته باشد:

- Index داده به ترتیب داده تست
- Offensive (پیش بینی باینری شما 0: غیر توهین آمیز و 1: توهین آمیز)

در لینک یک فایل نمونه قابل دانلود می‌باشد.

## موارد آپلود

در نهایت باید موارد زیر را در قالب یک فایل زیپ در کورسور آپلود نمایید:

- کد کامل پروژه
- گزارش کامل پروژه شامل توضیحات روش یا روش‌های استفاده شده برای پیش‌پردازش و مدل پیش‌بینی و تحلیل شما از نتایج بدست آمده

در صورتی که با روش‌های مختلفی دقت‌های مختلف بدست آوردید؛ تمامی آن‌ها را هم در کد قرار دهید و هم در گزارش خود بیان کنید

لازم به ذکر است که در صورت آپلود نشدن هر کدام از کد یا گزارش، نمره پروژه 0 لحاظ خواهد شد.

## نمره‌دهی پروژه

باتوجه به scoreboard نهایی و رتبه و دقت افراد، نمرات scale خواهد شد. هم‌چنین بخشی از نمره به پیاده‌سازی شما و گزارش تعلق می‌گیرد. انتظار می‌رود دانشجویان از روش‌های مختلف استفاده کرده و تحلیل خود از نتایج مختلفشان را گزارش کنند. تحلیل نتایج به صورت ضریب در نمره نهایی اثر خواهد گذاشت.

## تذکرات

۱. فرمت نامگذاری فایل تکلیف به صورت زیر باشد.  
[Final Project][student id][student name]
۲. در صورت مشاهده هرگونه تقلب نمره 0 برای پروژه در نظر گرفته خواهد شد.
۳. در صورت وجود هرگونه سوال یا ابهام، از راه تلگرام مشکل خود را مطرح کنید.

موفق باشید