

# 全球航运预抵识别

## 一、赛题背景

随着全球化的发展，海上运输在国际贸易中占据核心地位。为了优化物流管理和减少运输延误，准确预测船舶的到达时间和到达港口变得至关重要。本次挑战赛旨在通过分析船舶动态信息数据，提高港口到港口预测的准确性，从而助力全球物流行业的效率提升。

## 二、赛题任务

参赛者需要利用提供的船舶动态信息数据集，建立模型预测船舶的预计抵达港口（`end_port_code`）和预计抵达时间

（`leg_end_postime`）。选手需要自行进行数据预处理操作，应当充分利用提供数据集中各个字段的特征，并构建算法模型来预测船舶的预计抵达港口和预计抵达时间。

- 鼓励参赛者运用先进的数据分析、机器学习、深度学习等技术。

- 允许使用额外的公开数据源，以增强模型的准确性和泛化能力。

- 重视数据预处理和特征工程，以提高模型性能。

## 三、数据描述

1. 船舶动态信息数据训练集：包含历史船舶抵达各港口的时间记录，航行轨迹等关键参数。

2. 船舶动态信息数据测试集：仅包含船舶 MMSI、船舶起始港与船舶起始时间。

3.船舶静态信息数据：包含有关船舶的特征描述，如 MMSI 号、船舶类型、建造年份、载重吨、尺寸和吃水深度等信息。

4.船舶静态信息数据（船型映射关系表）：用于将船舶类型和子类型代码映射到相应的描述信息。

5.港口静态信息数据：包含有关港口的基本描述，如港口代码、名称、位置、国家、类型等信息。

#### 四、模型预测结果评分标准

选手需要同时预测船舶的预计抵达港口与预计抵达时间，即预测结果需包括每条船舶的预计抵达港口和预计抵达时间。当预计抵达港口预测错误时，该次动态的预测结果，该次动态的预测得分将得 0 分。当预计抵达港口预测正确时，将继续评估预计抵达时间与实际抵达时间的差值，单位为小时，如果预计抵达时间与真实抵达时间差值不超过 7 天，则计算(168 - 预测时间)与 168 的比值为该次动态的预测得分。

如果预计抵达时间与真实抵达时间差值超过 7 天，该次动态的预测得分将得 0 分。

预测结果评分标准公式为：

$$\text{评分} = \begin{cases} \frac{168 - \text{预测时间差值}}{168}, & \text{如果港口预测正确且时间差值} \leq 168 \text{ 小时} \\ 0, & \text{其他情况} \end{cases}$$

船舶动态信息数据测试集共有 16732 条数据，则该模型预测结果满分为 16732 分。

## 五、数据格式

船舶动态信息数据训练集共有 807 条船舶的 33968 条船舶动态数据，船舶动态信息数据训练集格式说明：

列名	类型	说明
ship_mmsi	str	船舶的 MMSI (Maritime Mobile Service Identity) 号码
start_port_code	str	起始港口的代码
end_port_code	str	目的港口的代码
leg_start_postime	str	航段开始的时间，格式为 yyyy-MM-dd HH:mm:ss+tz
leg_end_postime	str	航段结束的时间，格式为 yyyy-MM-dd HH:mm:ss+tz
route_line	str	航线的表示，格式为 WKT (Well-Known Text) 的 LINESTRING 类型
distance	float	航段距离，单位为公里或相似单位

船舶动态信息数据测试集共有 807 条船舶的 19191 条船舶动态数据，船舶动态数据测试集格式说明：

列名	类型	说明
uuid	str	唯一标识符，用于区分每条记录的独特 ID
ship_mmsi	str	脱敏后的船舶 MMSI 号，用于标识船舶的唯一号码
start_port_code	str	起始港口代码，表示货物出发的港口
leg_start_postime	datetime	起始航段的开始时间，格式为：yyyy-MM-dd HH:mm:ss+tz

end_port_code	str	需补充提交，目的港口代码，表示货物到达的港口
leg_end_postime	datetime	需补充提交，起始航段的结束时间，格式为：yyyy-MM-dd HH:mm:ss+tz

船舶静态信息数据包含 843 条船舶的静态信息，船舶静态信息数据格式说明：

列名	类型	说明
ship_mmsi	str	船舶的 MMSI 号
vessel_type	int	船舶类型代码
vessel_sub_type	int	船舶子类型代码
build_year	int	船舶建造年份
deadweight	int	船舶的载重吨
length	float	船舶的长度（米）
width	float	船舶的宽度（米）
height	float	船舶的高度（米）
draught	float	船舶的吃水深度（米）
max_speed	float	船舶的最大速度（节）

船舶静态信息数据（船型映射关系表）包含 6 类船舶的类型映射关系，与 41 个船舶子类型映射关系，船舶静态信息数据格式说明：

列名	类型	说明
dict_type	str	字典类型（字符串类型）
dict_code	str	字典代码（字符串类型）
parent_code	str	父级字典代码，可为空（字符串类型），

		仅当 dict_type 为 vessel_sub_type 时不为空
name_en	str	英文名称（字符串类型）
name_cn	str	中文名称（字符串类型）

港口静态信息数据包含 1101 个港口的地理位置数据，港口静态信息数据格式说明：

列名	类型	说明
port_code	str	港口代码（字符串类型）
ctry_code	str	国家代码（字符串类型）
name_en	str	英文名称（字符串类型）
name_cn	str	中文名称（字符串类型）
lon	float	经度（浮点数类型）
lat	float	纬度（浮点数类型）
timezone_offset	int	时区偏移（整数类型，单位为小时）