

# 推荐模型离线评估

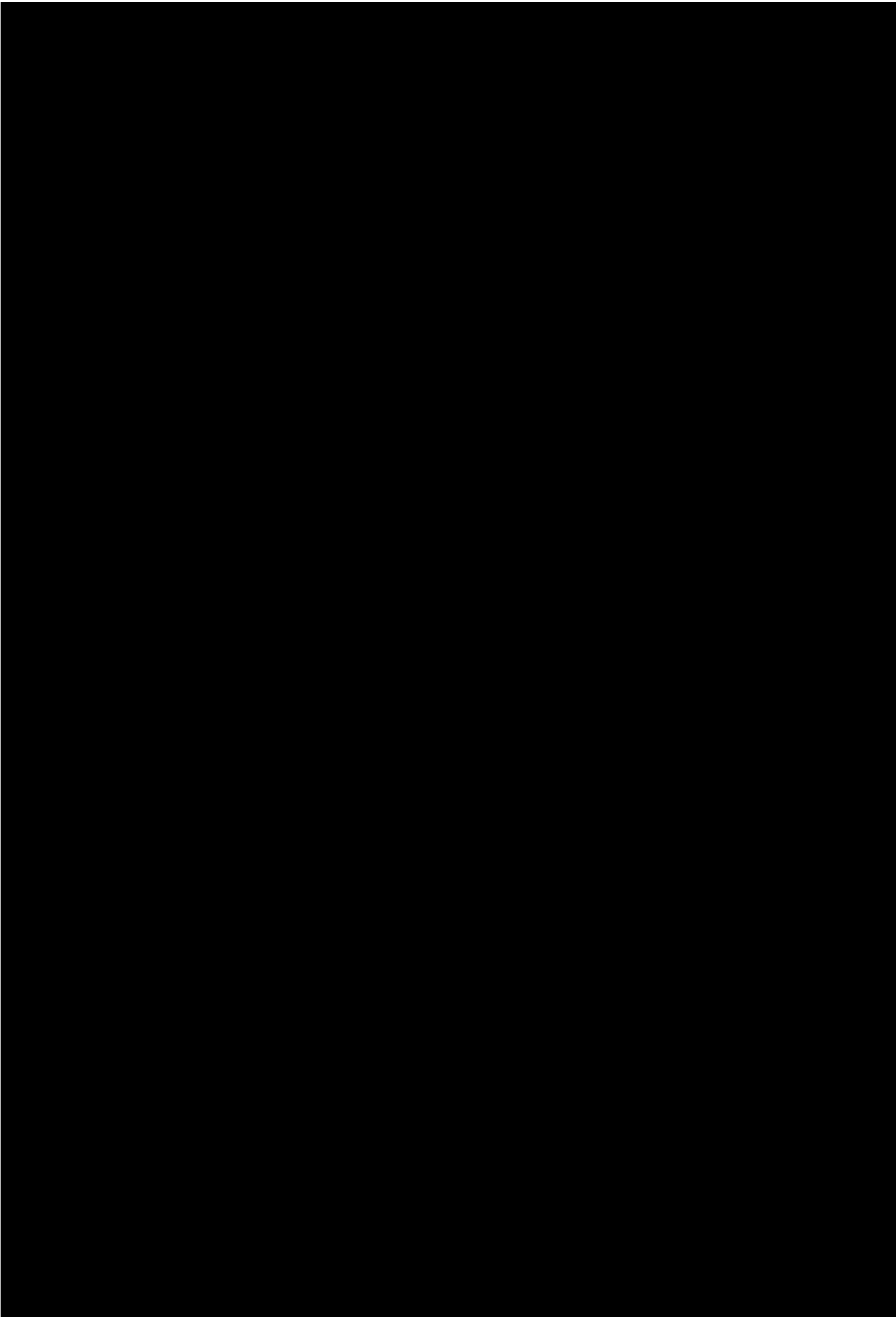
在之前的课程中我们学习了如何搭建推荐系统。在有了推荐系统之后我们下一步面临的问题往往就是如何知道系统的表现好坏？刚刚搭建完一个系统后我们显然不能立即将其上线，因为整个模型的效果是未知的，贸然就上线往往可能会造成不可预计的后果。因此使用合适的方法对模型进行离线评估就非常重要了。

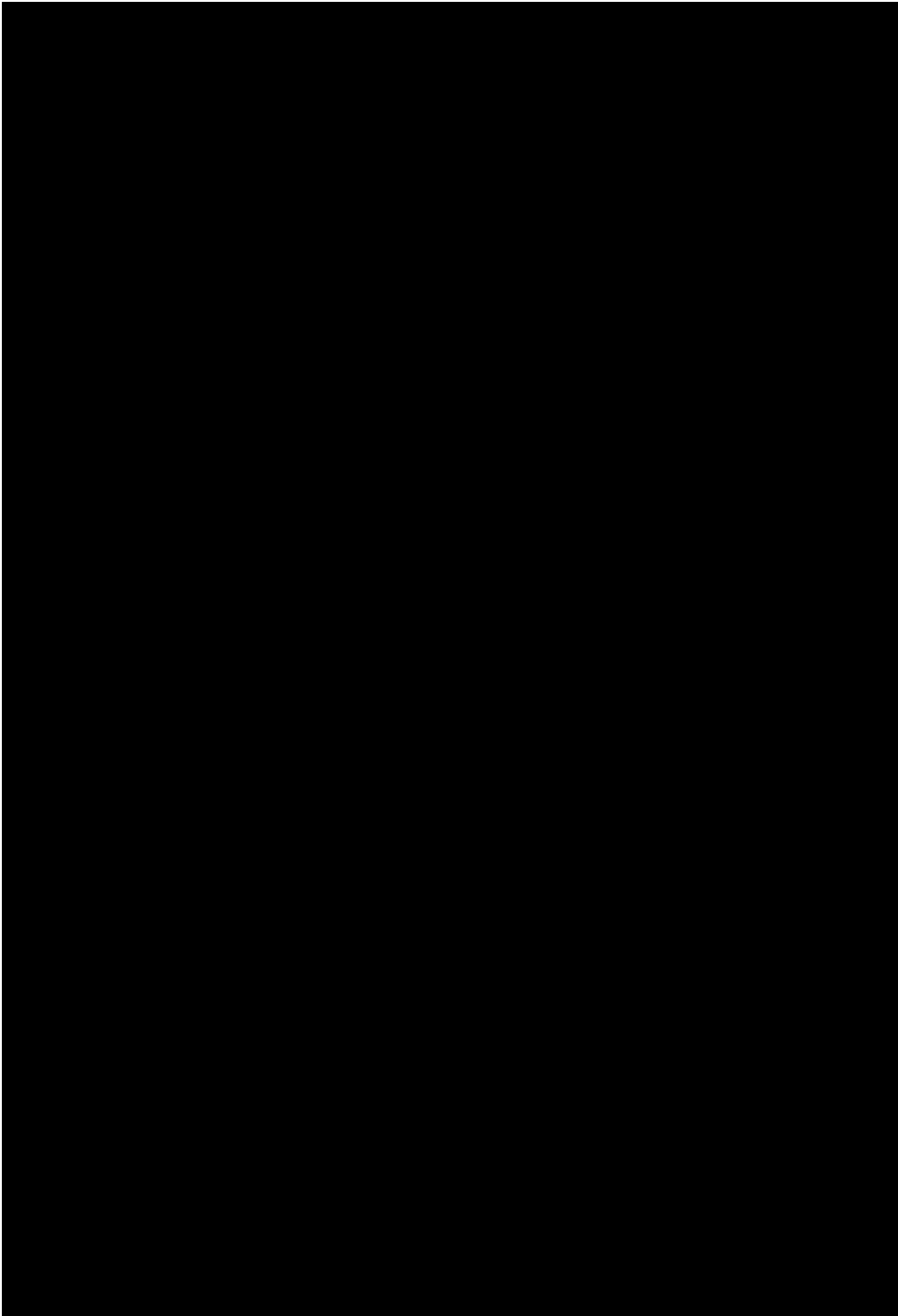
在本节中我们将讨论几个业界十分常用的离线评估指标，并对比各自的优劣。

## RMSE均方根误差

对于任何一个回归模型的评判，最容易想到的指标就是模型的误差。也就是用模型的预测值和真实值进行比较，然后看一下误差是多还是少。而在这类算法中最经典的就是RMSE均方根误差，该算法如下：

$$RMSE = \sqrt{\frac{\sum (y_i - \widehat{y}_i)^2}{|T|}}$$
 RMSE= | T |







$$\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

其中 $y_i$ 为模型预测值， $\hat{y}_i$ 为真实样本值。  
RMSE虽然很常用，但是它也有自己的缺点，那就是对异常值比较敏感。当用它来评估如停留时长等没有明确范围的指标的时候可靠度就会下降。

### 准确、召回率

对于推荐系统这类的二分类问题而言，准确、召回率也是非常常用的评估指标。

其中准确率被定义为：

$$\text{Accuracy} = \frac{\text{用户感兴趣的物品}}{\text{推荐物品}}$$

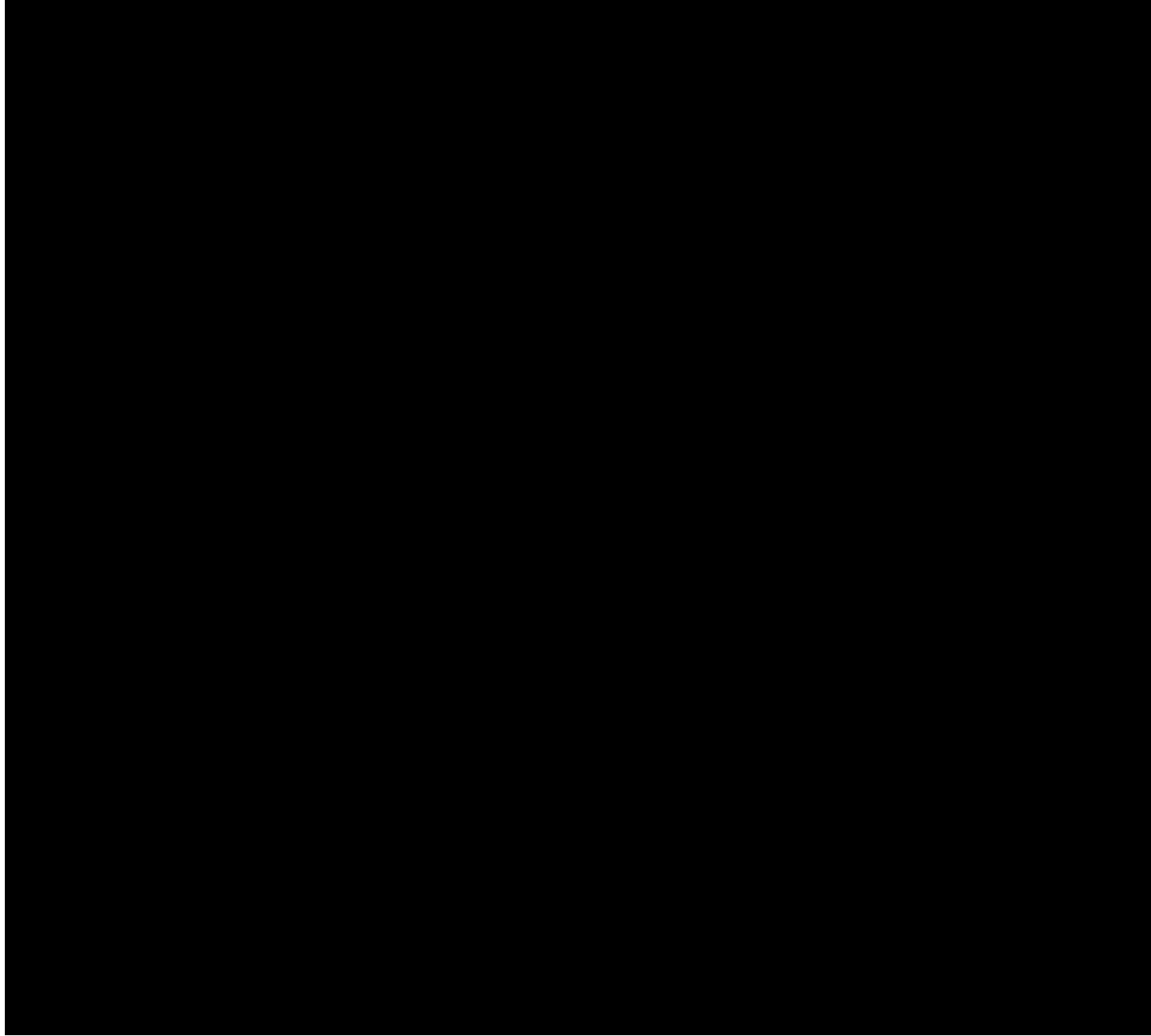
准确率直观的理解就是：在推荐的物品中，有多少是用户真正感兴趣的。

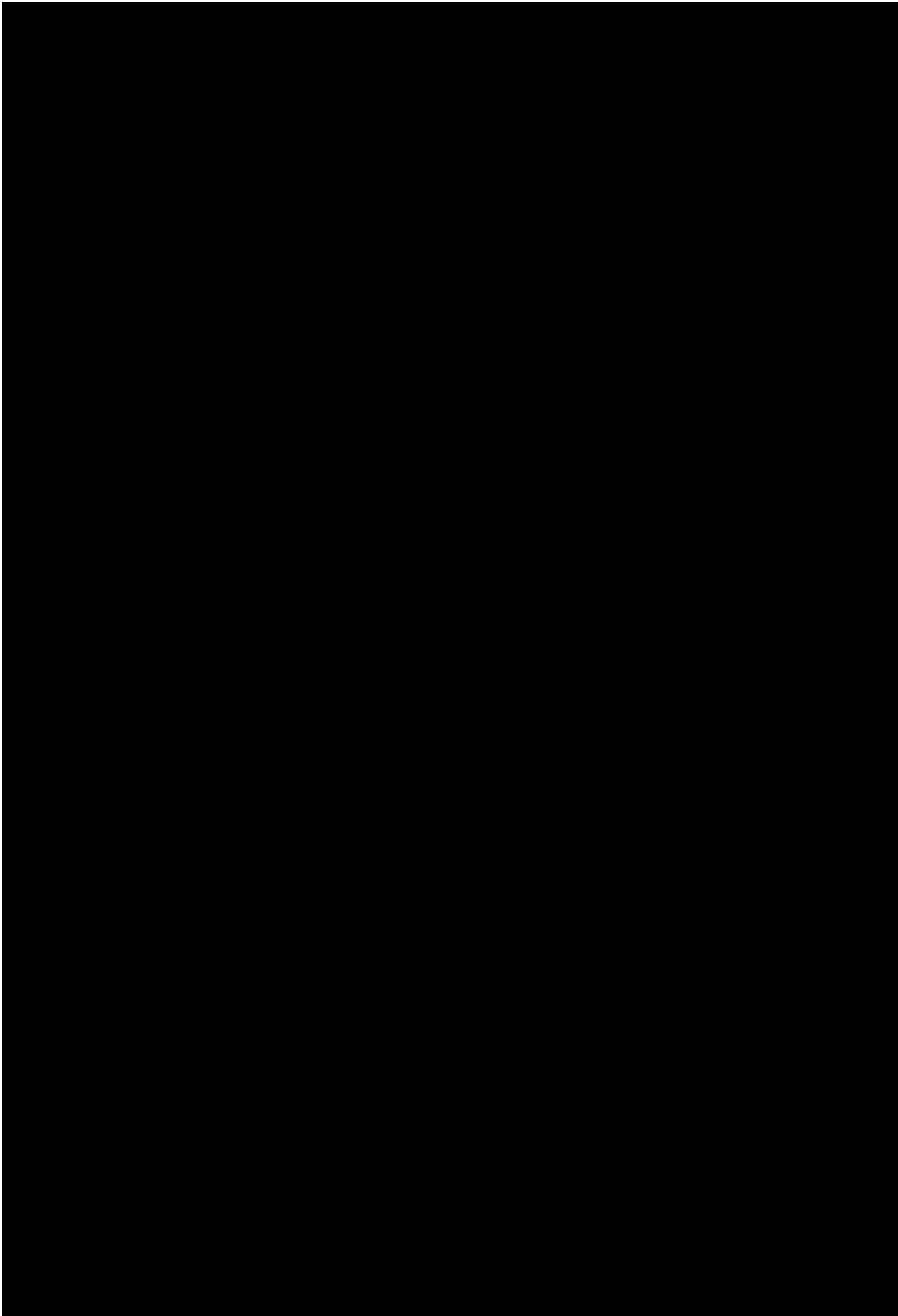
相对的，召回率的定义为：

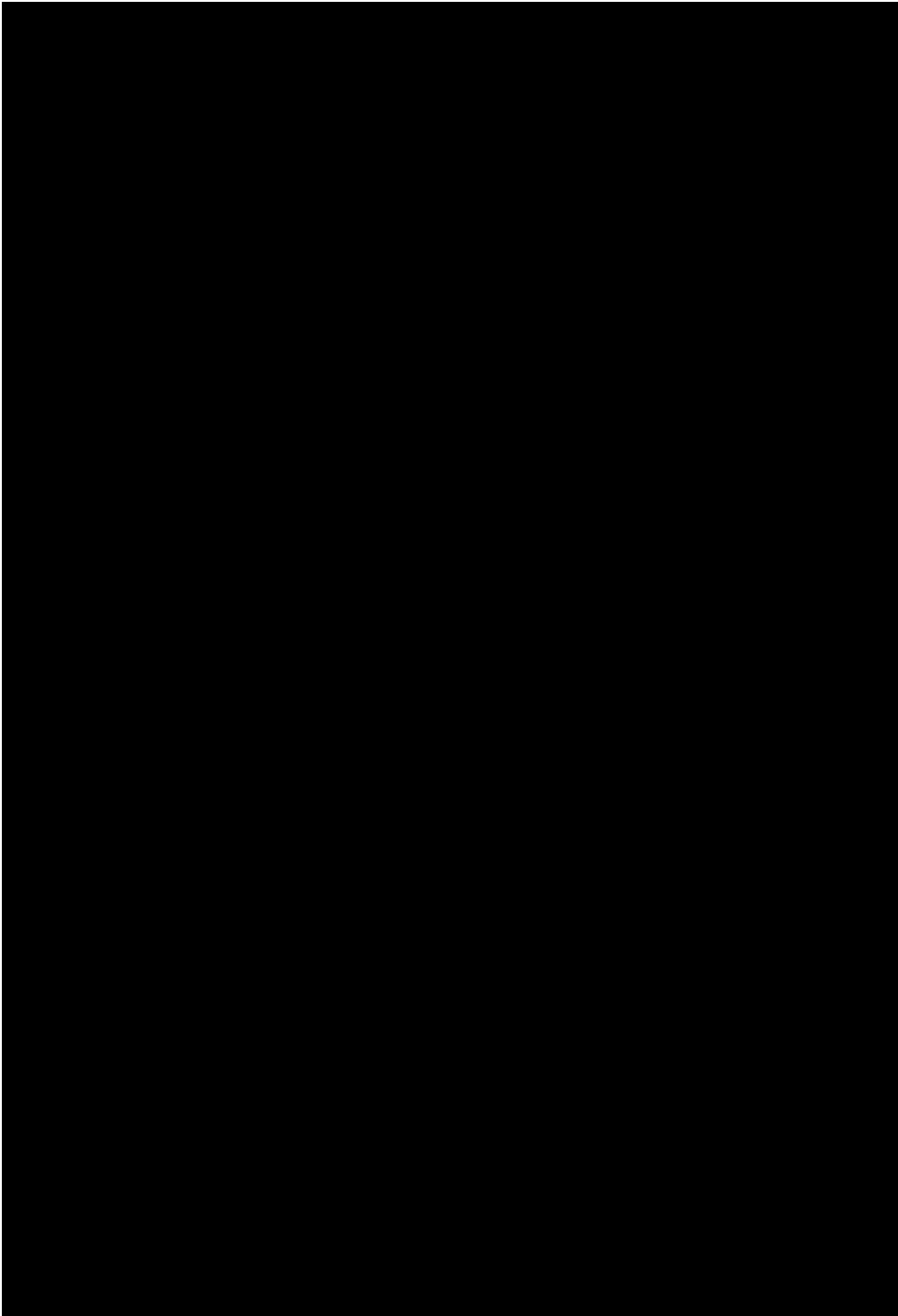
$$\text{Recall} = \frac{\text{用户感兴趣的物品}}{\text{所有用户感兴趣的物品}}$$

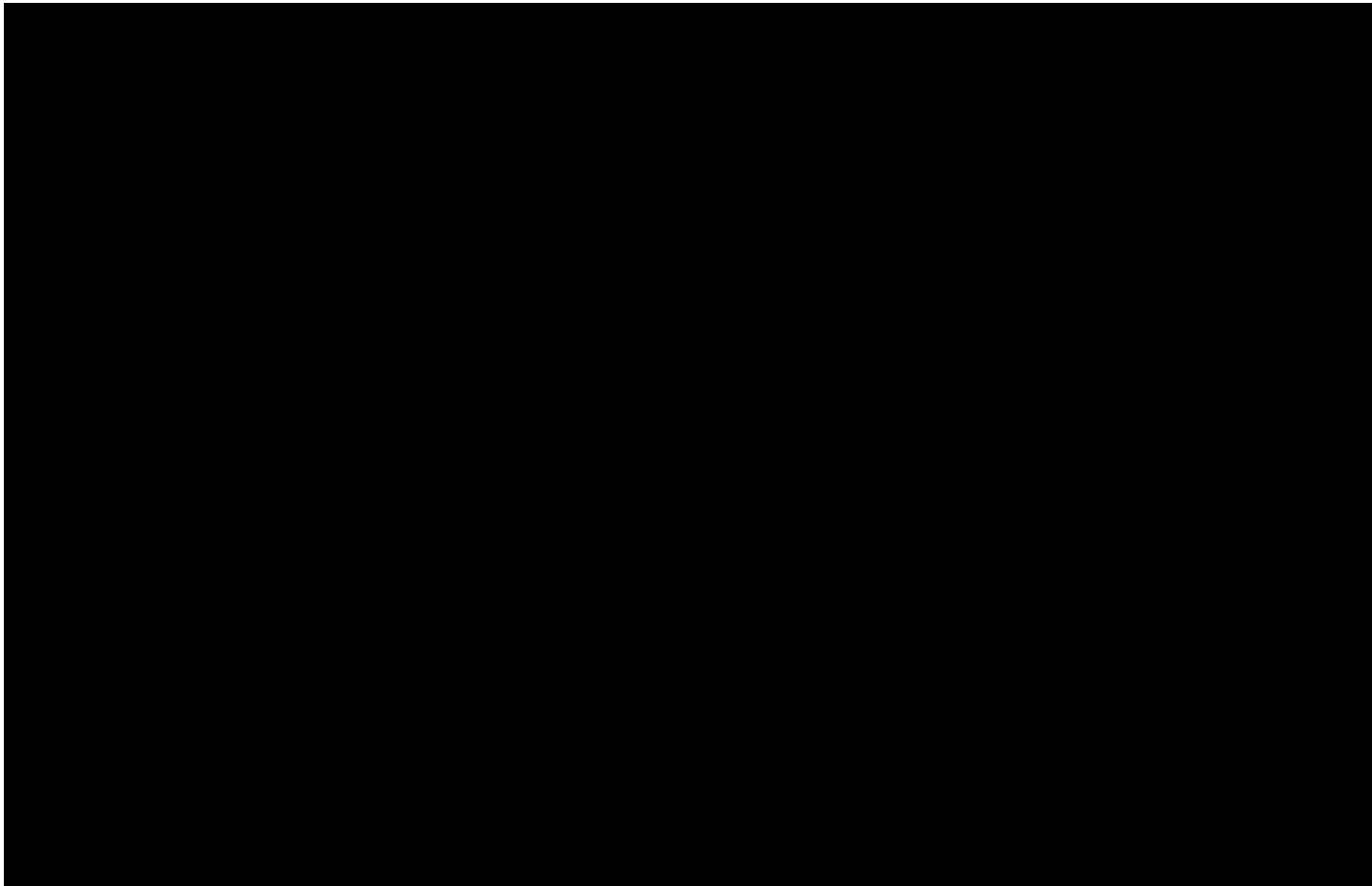
召回率直观的理解就是：在所有用户感兴趣的物品中，系统找出了多少。

准确率和召回率是两个相辅相成的指标，不应该只强调其中的一个。为了直观的同时观察这两个指标，我们可以引入F值，F值等于准确率和召回率的调和平均数，即：  
$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$









precision\*recall

总结：有了以上几种指标，我们就可以初步对一个模型的好坏进行离线评估了，如果一个模型在绝大部分指标上都要表现的更优秀，那么我们对该模型的实际表现就能有较大的信心。但是模型上线后是否真的有效还需要进行其他更细致的评估以及在线AB测试才能得到最终结论。