

אלגוריתם ID3

נתוני שידור שירים ברדיו:

שודר?	שיחד שדרן	זמר/להקה	ז'אנר	אורך השיר
כן	כן	זמר	מזרחי	עד 2 דקות
לא	לא	להקה	מזרחי	עד 2 דקות
כן	כן	להקה	רוק	מעל 5 דקות
לא	לא	זמר	רוק	מעל 5 דקות
לא	לא	להקה	פופ	מעל 5 דקות
לא	לא	זמר	מזרחי	מעל 5 דקות
כן	לא	זמר	פופ	עד 2 דקות
לא	לא	להקה	פופ	עד 2 דקות
כן	לא	זמר	מזרחי	עד 2 דקות
לא	לא	להקה	רוק	עד 2 דקות

שודר?	שיחד שדרן	זמר/להקה	ז'אנר	אורך השיר
כן	לא	זמר	רוק	עד 2 דקות
כן	כן	להקה	רוק	2 - 5 דקות
כן	כן	להקה	מזרחי	2 - 5 דקות
כן	לא	זמר	מזרחי	2 - 5 דקות
כן	לא	להקה	פופ	2 - 5 דקות
כן	לא	זמר	פופ	2 - 5 דקות
לא	לא	להקה	רוק	2 - 5 דקות
לא	לא	זמר	רוק	2 - 5 דקות

חישוב אנטרופיה - אורך השיר

$$I_{<2} = -\frac{4}{7} \log_2 \frac{4}{7} - \frac{3}{7} \log_2 \frac{3}{7} = 0.98523$$

$$I_{2-5} = -\frac{2}{7} \log_2 \frac{2}{7} - \frac{5}{7} \log_2 \frac{5}{7} = 0.86312$$

$$I_{>5} = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0.81128$$

• אנטרופיה ממוצעת:

$$I = \frac{7}{18} \cdot 0.98523 + \frac{7}{18} \cdot 0.86312 + \frac{4}{18} \cdot 0.81128 = 0.89908$$

חישוב אנטרופיה - ז'אנר

$$I_{miz} = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0.9183$$

$$I_{rock} = -\frac{3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7} = 0.98523$$

$$I_{pop} = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97095$$

• אנטרופיה ממוצעת:

$$I = \frac{6}{18} \cdot 0.9183 + \frac{7}{18} \cdot 0.98523 + \frac{5}{18} \cdot 0.97095 = 0.95895$$

חישוב אנטרופיה - זמר/להקה

$$I_{singer} = -\frac{6}{9} \log_2 \frac{6}{9} - \frac{3}{9} \log_2 \frac{3}{9} = 0.91830$$

$$I_{group} = -\frac{4}{9} \log_2 \frac{4}{9} - \frac{5}{9} \log_2 \frac{5}{9} = 0.99108$$

• אנטרופיה ממוצעת:

$$I = \frac{9}{18} \cdot 0.91830 + \frac{9}{18} \cdot 0.99108 = 0.95469$$

חישוב אנטרופיה - שוחד

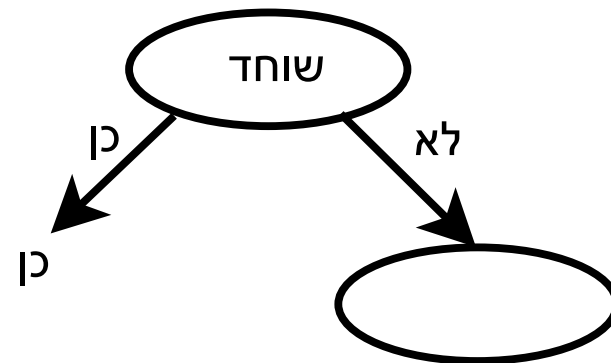
$$I_{bribe} = -\frac{4}{4} \log_2 \frac{4}{4} = 0$$

$$I_{nobribe} = -\frac{6}{14} \log_2 \frac{6}{14} - \frac{8}{14} \log_2 \frac{8}{14} = 0.98523$$

• אנטרופיה ממוצעת:

$$I = \frac{4}{18} \cdot 0 + \frac{14}{18} \cdot 0.98523 = 0.76629$$

שלב ראשון



- נבחר באפשרות השוחד, משום שזו האפשרות עם האנטרופיה הקטנה ביותר.
- עבור שוחד=כן, סיימנו. נמשיך עבור שוחד=לא.

שלב 2: חישוב אנטרופיה - אורך השיר

$$I_{<2} = -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6} = 1$$

$$I_{2-5} = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97095$$

$$I_{>5} = -\frac{3}{3} \log_2 \frac{3}{3} = 0$$

• אנטרופיה ממוצעת:

$$I = \frac{6}{14} \cdot 1 + \frac{5}{14} \cdot 0.97095 + \frac{3}{14} \cdot 0 = 0.77534$$

שלב 2: חישוב אנטרופיה - ז'אנר

$$I_{miz} = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1$$

$$I_{rock} = -\frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} = 0.72193$$

$$I_{pop} = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97095$$

• אנטרופיה ממוצעת:

$$I = \frac{4}{14} \cdot 1 + \frac{5}{14} \cdot 0.72193 + \frac{5}{14} \cdot 0.97095 = 0.89031$$

שלב 2: חישוב אנטרופיה - זמר/להקה

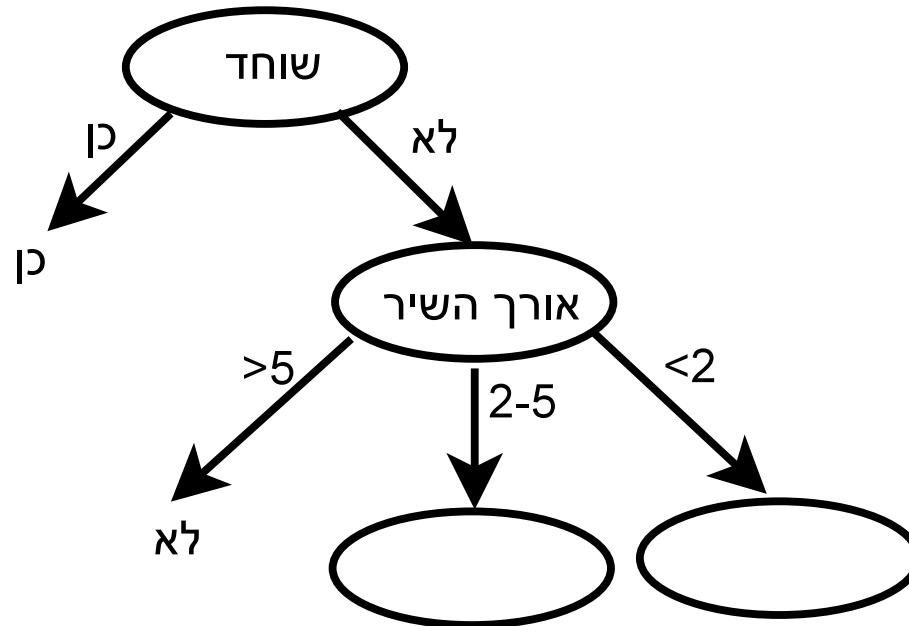
$$I_{singer} = -\frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} = 0.95443$$

$$I_{group} = -\frac{1}{6} \log_2 \frac{1}{6} - \frac{5}{6} \log_2 \frac{5}{6} = 0.65002$$

• אנטרופיה ממוצעת:

$$I = \frac{8}{14} \cdot 0.95443 + \frac{6}{14} \cdot 0.65002 = 0.82397$$

שלב שני



- נבחר באפשרות של אורך השיר, משום שזו האפשרות עם האנטרופיה הקטנה ביותר.
- עבור שירים באורך יותר מ- 5 דקות סיימנו.

שלב 3, אורך שיר > 2 : חישוב אנטרופיה - ז'אנר

$$I_{miz} = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$I_{rock} = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$I_{pop} = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

• אנטרופיה ממוצעת:

$$I = \frac{2}{6} \cdot 1 + \frac{2}{6} \cdot 1 + \frac{2}{6} \cdot 1 = 1$$

שלב 3, אורך שיר > 2 : חישוב אנטרופיה - זמר/להקה

$$I_{singer} = -\frac{3}{3} \log_2 \frac{3}{3} = 0$$

$$I_{group} = -\frac{3}{3} \log_2 \frac{3}{3} = 0$$

• אנטרופיה ממוצעת:

$$I = \frac{3}{6} \cdot 0 + \frac{3}{6} \cdot 0 = 0$$

• קיבלנו הבחנה מלאה בין הנתונים.

שלב 3, אורך שיר 5-2: חישוב אנטרופיה - ז'אנר

$$I_{miz} = -\frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$I_{rock} = -\frac{2}{2} \log_2 \frac{2}{2} = 0$$

$$I_{pop} = -\frac{2}{2} \log_2 \frac{2}{2} = 0$$

• אנטרופיה ממוצעת:

$$I = \frac{1}{5} \cdot 0 + \frac{2}{5} \cdot 0 + \frac{2}{5} \cdot 0 = 0$$

• קיבלנו הבחנה מלאה בין הנתונים.

סיכום

