

DASAR ANALISA DATA DENGAN MENGGUNAKAN TEKNIK PENGENALAN DAN KLASIFIKASI POLA*)

Oleh: Ir. RIDUAN NAZIR**)

RINGKASAN

Dalam tulisan ini akan dibicarakan dasar analisa data dengan teknik statistik pengenalan pola maupun cara mengatur pola. Pembicaraan ditekankan pada bagaimana mengolah data remote sensing dari data multispectral untuk memperoleh informasi sumber-sumber daya alam (Bumi).

Modal umum pada pengenalan pola yang dipilih memerlukan Feature selection dan Extraction, pemilihan suatu fungsi diskriminan. Dalam hal ini yang dibicarakan adalah mengenai Feature Principle dan pemilihan suatu fungsi diskriminan.

Pembicaraan meliputi hal-hal antara lain; teori Bayesian, fungsi diskriminan kemiripan terbesar (maximum likelihood discriminant function), pengolahan data yang dibimbing (supervised processing), unsupervised processing (non supervised processing).

1. PENDAHULUAN

Dengan menggunakan Multispectral scanner dan Multiband photography memungkinkan pengumpulan data dari daerah daratan dan laut dalam bentuk signal yang sesuai dengan pengolahan data dalam digital computer. Banyaknya informasi yang diperoleh dari imageri multi-spectral akan mempersulit interpretasi visual dan akan memerlukan waktu untuk menginterpretasikannya.

Dengan menggunakan digital computer maka data imageri multi-spectral itu dapat diolah dengan pertolongan teknik-teknik pengenalan pola, karena teknik ini dapat menyediakan cara kerja sendiri (automatic procedure) untuk menentukan tingkatan setiap elemen permukaan bumi yang direkam atau pixel, dimana bentuk pixel ini sebagai hal yang klasifikasinya benar (a correct classification).

Pembicaraan akan menguraikan pula teori dasar pengenalan pola dan cara mengatur pola secara berurut yang dipakai dalam digital image processing (pengolahan data secara digital). Akan diutarakan juga hal-hal lain seperti teori Bayesian, fungsi diskriminan kemiripan terbesar, pengolahan data yang dibimbing (supervised processing), pengolahan data tak dibimbing (un supervised processing). Dalam pengolahan data yang dibimbing mencakupi metoda statistik dan metoda diskriminan. Sedangkan dalam pengolahan data tak dibimbing mencakupi metoda subgroup-(clustering). Untuk pengolahan data yang dibimbing dan pengolahan data tak dibimbing mempunyai keuntungan yang nyata bagi masing-masing metoda.

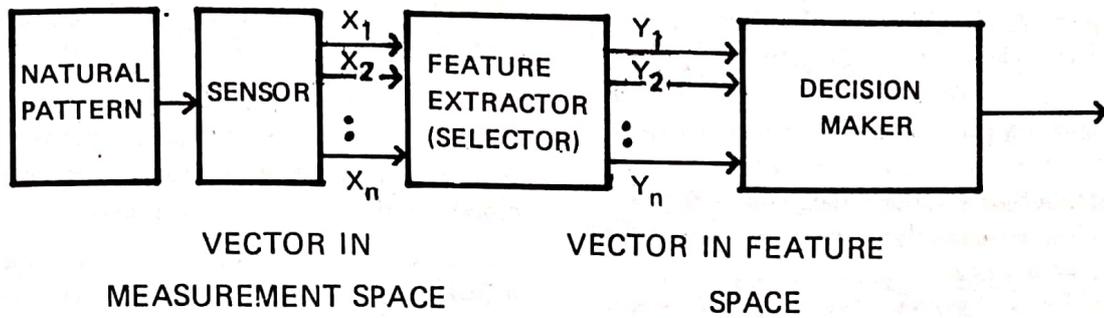
2. DASAR TEKNIK PENGENALAN POLA DAN KLASIFIKASI POLA

2.1. Model Pengenalan Pola

Pada gambar di bawah ini diperlihatkan sistem umum pengenalan pola yaitu:

* Majalah LAPAN No. 16 Tahun ke IV

** Staf Proyek TELSA-LAPAN



Gambar 2.1. Model Pengenalan Pola.

Untuk mengumpulkan data dapat digunakan sensor-sensor antara lain; Multispectral Scanner, Multiband photography. Setiap elemen permukaan bumi yang direkam oleh sensor, dan sensor-sensor melaksanakan pengukuran sesuai dengan n band atau n channels. Sedangkan n pengukuran membatasi suatu titik dalam n dimensi Euclides space (pengukuran ruang), misalnya sistem yang menggunakan 2 detektor yang berbeda band (channel) maka sistem detektor ini mengadakan 2 pengukuran band bagi satu obyek, dan hal ini berarti ada 2 vector ruang yang ditunjukkan oleh 2 pengukuran dari 2 detektor yang berbeda band (channel) maka resultante kedua vector adalah obyek yang dideteksi yang ditunjukkan dalam bentuk golongan-golongan (classis). Setiap n pengukuran dapat ditunjukkan dan dinyatakan sebagai vector kolom dalam matriks berdimensi n;

$$x = (x_1, x_2, \dots, x_n) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

The Feature Extractor memilih dan meringkaskan vector pengukuran menjadi vector feature, dan dalam hal ini berarti terjadi transformasi dari ruang vector X_n ke ruang vector Y_m dimana m lebih kecil dari pada n.

Dimisalkan transformasi linear adalah; transformasi $Y = T X$, dimana transformasi T berbentuk matrices $n \times n$ dan transformasi T bukan correspondensi satu ke satu. Maka persoalannya adalah mencari matriks transformasi T sehingga diperoleh sebahagian besar informasi dalam vector ruang X_n berada dalam vector ruang Y_m , dan selanjutnya untuk menyelesaikan jawab transformasi itu dapat dipakai beberapa metoda (dalam hal ini tidak dibicarakan), misalnya metoda Gauss.

The Feature Selector memilih band (channel) yang paling baik dan memisahkan golongan yang satu terhadap yang lain. Dalam hal ini ada dua metoda untuk mencari jarak yaitu:

- * Mencari jarak rata-rata antara titik data terhadap titik data lainnya dalam satu golongan (class) atau dinamakan jarak intra golongan (intrakelas).
- * Mencari jarak rata-rata antara semua titik pada semua golongan (one class) terhadap setiap titik pada semua golongan lain atau dinamakan jarak antar golongan (antar kelas).

Untuk memperoleh band (channel) yang paling baik maka dipakai kriteria yang dinamakan figure Of Merit (FOM). Kriteria FOM adalah $Exp(-interclass\ distance/between\ twoclasses\ distance)$. Penggunaan yang lebih detail tentang FOM ini tidak dibicarakan dalam tulisan ini.

The Decision Maker mengerjakan/memilih suatu keputusan pada vector feature dari titik data yang diketahui kedalam golongan khusus (a particular class).

Dalam pelaksanaan tahap penggolongan (klasifikasi) dipakai beberapa urutan metoda antara lain; metoda-statistik, metoda aturan keputusan Bayesian, metoda aturan keputusan kemiripan terbesar.

2.2. Teknik Pengenalan Pola.

Untuk menandai pola dalam teknik pengenalan, digunakan dua metoda, yaitu:

- * Metoda pengolahan data yang dibimbing, dimana pelaksanaannya memerlukan dua ciri yaitu;
 - a) Sifat-sifat golongan (classes) yang berbeda yang merupakan ciri-ciri khusus (signatures).
 - b) Membandingkan feature dari pixel-pixel yang tak diketahui terhadap ciri-ciri standard.

* Metoda pengolahan data tak dibimbing, hal ini bertitik tolak pada usaha agar golongan-golongan yang berbeda membentuk kelompok terpisah dalam ruangan feature. Pelaksanaan penggroup-an disebut clustering. Tetapi hal ini masih sangat berguna dalam persoalan pengenalan pola. Pengerjaan clustering digunakan dua metoda utama yaitu:

- Cara bekerja berulang (the iterative procedures).
- Cara bekerja satu kali selesai (the one-pass procedurs).

3. SUPERVISED PROCESSING

Pengolahan data yang dibimbing ini bertujuan memisahkan satu golongan dari suatu obyek (one class of the obyek) terhadap golongan lainnya yang memerlukan asumsi-asumsi yang perlu diketahui sebagai ciri-ciri atau sifat golongan itu (the knowledge of some set of attributes known as signatures). Dalam hal ini fungsi diskriminan menggunakan ciri-ciri atau sifat-sifat tersebut untuk tujuan-tujuan penggolongan (clasification purposes).

Jadi dapat disimpulkan bahwa pengolahan ini memerlukan dua ciri-ciri yang perlu diketahui antara lain:

- * Sifat-sifat golongan yang berbeda merupakan ciri-ciri khusus.
- * Membandingkan feature dari pixel-pixel yang tidak diketahui terhadap ciri-ciri standard.

4. METODA STATISTIK

Dalam metoda ini (statistik approach) sering digunakan metoda fungsi kerapatan kemungkinan yang berhubungan dengan golongan-golongan pola (the pattern classes). Teknik statistik ini selalu tidak diketahui tentang hal-hal yang berhubungan dengan golongan pola dan hal ini hanyalah berdasarkan pertimbangan atau anggapan melalui training samples.

Pertimbangan-pertimbangan itu (anggapan-anggapan) dapat diperoleh dengan dua cara yaitu:

- * Pertimbangan untuk mencari parameter mean, standard.deviasi, kovarian dan lain-lain.
- * Pertimbangan untuk mencari non parametrik, dimana bentuk fungsi kerapatan yang tidak diketahui umumnya sulit untuk melaksanakannya. Dan hal ini selanjutnya tidak dibicarakan.

Dengan metoda statistik golongan-golongan pola dihubungkan dengan kemungkinannya.

Misalnya: $p(W_i)$, $i = 1, 2, \dots, n$ dimana $p(w_i)$ adalah kemungkinan dari setiap golongan pola (the probability of each pattern class). Dan misalkan pula bahwa:

$$p(x/w_i) = i = 1, 2, \dots, n$$

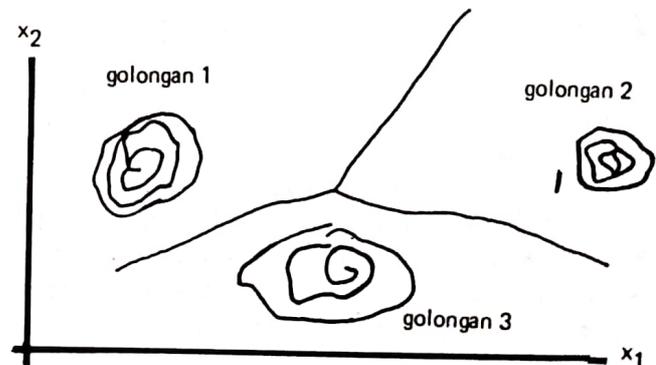
adalah kemungkinan keadaan dari suatu vector feature x kepunyaan golongan w_i (the conditional probability of a feature vector x belongs to a class w_i). Untuk dasar informasi dipakai: $p(w_i)$ dan $p(x/w_i)$ dimana $j = 1, 2, \dots, n$

5. FUNGSI DISKRIMINAN

Pola-pola yang menggambarkan suatu golongan khusus memperlihatkan beberapa macam perubahan yang berhubungan dengan sifat-sifat benda alam (natural thing), karena vector-vector-nya dari pola yang sama (satu pola) cenderung untuk membentuk suatu liputan titik-titik (The vector from the same pattern tend to form a cloud of points).

Untuk jelasnya dapat dilihat dalam gambar berikut ini:

Gambar di bawah ini menunjukkan daerah-daerah penentuan (keputusan) dan liputan titik-titik.



Pekerjaan mengatur pola adalah membagi ruang feature menjadi daerah-daerah penentuan (keputusan) dimana setiap daerah bersesuaian dengan suatu golongan khusus. Untuk setiap titik data yang berada dalam suatu daerah klasifikasi khusus ditunjukkan sebagai suatu golongan yang berhubungan dengan daerah itu dan daerah yang

telah diklassifikasikan diketahui sebagai daerah klasifikasi tertentu. Jadi merencanakan suatu pengaturan pola tujuan akhir adalah membimbing beberapa cara bekerja untuk menentukan kelas-kelas tertentu (keputusan) sehingga mencapai kriteria klasifikasi.

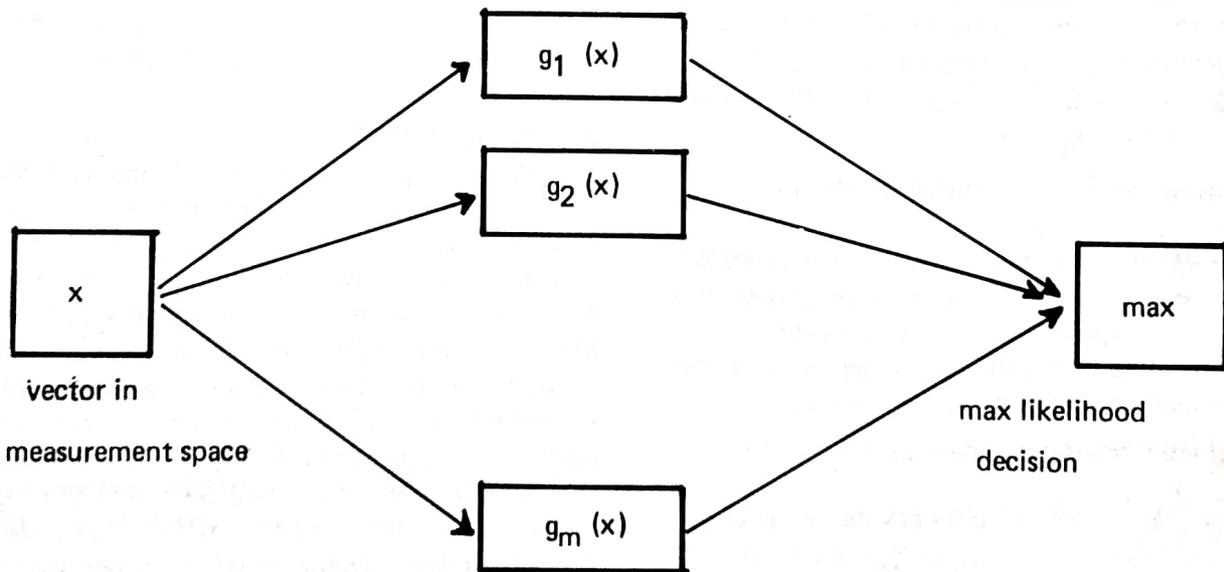
Dalam pemilihan kriteria itu maka dapat ditunjukkan suatu bentuk matematik yang dinamakan sebagai fungsi diskriminan, dan dianggap bahwa ada m golongan-golongan pola serta fungsi diskriminannya adalah ; $g_1(x), g_2(x) \dots g_j(x)$ dimana fungsi diskriminan itu bernilai sebagai fungsi x sehingga ;

$$g_i(x) \text{ lebih besar daripada } g_j(x)$$

untuk semua harga-harga x dalam daerah yang berhubungan dengan golongan ke i ($i \neq j$). Untuk kontinuitas fungsi diskriminan pada garis-garis batas penentuan ;

$$g_i(x) - g_j(x) = 0.$$

Dalam metoda untuk pengaturan pola (pattern classifier) dibutuhkan fungsi diskriminan yang optimum. Umumnya garis-garis batas penentuan yang optimum bentuknya linear-Gambar pengaturan pola dapat dilihat dibawah ini.



Gambar – 3. Pengaturan Pola (Pattern Classifier).

6. ATURAN PENENTUAN BAYESIAN

Aturan penentuan yang optimal bertujuan mengurangi kehilangan rata-rata yang disebut aturan penentuan - Bayesian (Bayesian decision rule). Sesuai dengan aturan ini maka suatu vector feature x kepunyaan golongan w_i memenuhi ;

$$p(x/w_i) p(w_i) = p(x/w_j) p(w_j) \dots \dots i \neq j$$

dimana :

$p(w_i)$ adalah kemungkinan atau (an apriori function).

$p(x/w_j)$ adalah keadaan fungsi kerapatan kemungkinan atau (a conditional probability density function).

Selanjutnya fungsi kerapatan kemungkinan disederhanakan dengan anggapan bahwa $p(x/w_i)$ sebagai fungsi multivariate kerapatan Gaussian (a multivariate Gaussian density function). Dengan menggunakan metoda parametrik dan anggapan bahwa suatu golongan (a class) mempunyai keadaan fungsi kerapatan kemungkinan maka fungsi multivariate kerapatan Gaussian menjadi ;

$$p(x/w_i) = \frac{1}{(2\pi)^{n/2} |C_i|^{1/2}} \exp [-\frac{1}{2} (x-U_i)^T C_i^{-1} (x-U_i)]$$

- dimana : U_j adalah vector pengharapan (expectation vector)
 C_j adalah matrik kovarian (covariance matrix).
 T adalah Transpose $i = 1, 2, \dots, n$
 n adalah jumlah pengukuran atau band atau channel.

Untuk vector pengharapan (mean) U_j dari n pengukuran kepunyaan golongan w_j maka dapat dianggap dengan vector arithmetik mean sebagai berikut :

$$U_j = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix}, \quad m_j = \frac{1}{n_t} \sum_{i=1}^{n_t} x_{ji}, \quad \begin{matrix} j = 1, \dots, n \\ i = 1, \dots, n_t \end{matrix}$$

Sedangkan kovariance dari n pengukuran kepunyaan golongan w_j dapat dianggap sebagai berikut :

$$C = (s_{ij}); \quad s_{ij} = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (x_{ji} - m_j)(x_{ki} - m_k)$$

$j = 1, \dots, n$
 $k = 1, \dots, n$

dimana, n_t adalah jumlah pixel-pixel dalam training sample.

7. ATURAN PENENTUAN KEMIRIPAN TERBESAR

Aturan penentuan kemiripan terbesar (the maximum likelihood decision rule) menunjukkan suatu vector feature x yang diberikan pada pola golongan w_j yang diberikan, adalah lebih besar dari pada untuk w_j

$$\text{maka, } p(w_j) = \dots = p(w_m) = 1, \quad i = 1, \dots, \frac{1}{m}$$

dan fungsi $p(x/w_j)$ menjadi suatu fungsi dalam aturan penentuan kemiripan terbesar (maximum likelihood decision rule). Jadi suatu vector feature x termasuk dalam golongan w_j jika memenuhi keadaan-keadaan sebagai berikut :

$$p(x/w_j) \text{ lebih besar daripada } L p(x/w_i), \quad i \neq j$$

$$\frac{1}{2}(x - U_j)^T \sum_i^{-1} (x - U_i) \text{ lebih kecil dari pada } D.$$

dimana, L adalah perbandingan kemiripan (likelihood ratio)

D adalah jarak ambang (threshold distance).

8. UNSUPERVISED PROCESSING

Pengolahan data tak dibimbing (unsupervised processing) mempunyai perbedaan utama dengan pengolahan data yang dibimbing (supervised processing) dalam hal-hal dimana unsupervised processing tidak memerlukan training sample yang ditandai (ditentukan jenis-jenisnya).

Dalam pelaksanaannya berdasarkan pada usaha-usaha agar pada golongan-golongan yang berbeda membentuk group-group dalam ruang feature. Ada beberapa hal yang menjadi alasan mengapa unsupervised processing dilaksanakan antara lain :

- * Pengumpulan training sample yang ditandai akan memerlukan waktu (banyak waktu terbuang) meskipun memberikan hasil yang sangat berharga.
- * Pemanfaatan sifat-sifat dari golongan pola sangat lambat pemakaiannya dan tergantung waktu.
- * Memulai suatu penyelidikan (investigation) data yang berhubungan dengan adalah obyek dengan studi radiasi elektromagnetik dari obyek dan susunan datanya.

Dalam menjawab pertanyaan-pertanyaan itu akan membawa kita kepada berbagai usaha untuk mem-formulasi pengaturan pola-pola (patterns classification) misalnya dengan cara pelaksanaan peng-groupan atau clustering.

9. CLUSTERING

Peng-group-an data melalui golongan-golongan yang tak ditandai ciri-cirinya (wheat, forest, sugarcane, corn, soil, chick pea, pigeon pea dan lain-lain) tetapi melalui hubungannya dengan nature yang berpautan dengan set titik-titik data. Misalnya kita ingin menginterpretasikan suatu daerah hutan tanpa memandang usia daerah hutan itu, tetapi yang dilihat adalah warna-warna hutan pada umumnya berwarna hijau. Dan kemudian dari data ground truth yang telah diketahui mengenai warna hijau adalah hutan, maka dapat memperkirakan berapa luasnya hutan itu dan berapa banyaknya pohon rata-rata. Walaupun cara kerja penggroup-an (clustering) tidak mempunyai arti theoretical properties, tetapi masih sangat berguna dalam persoalan-persoalan

pengenalan pola. Diambil suatu contoh lain, misalnya ; menginterpretasikan daerah corn (atau wheat atau the other trees) maka terlebih dahulu harus dikenal bahwa daerah sekitarnya banyak mengandung pohon-pohon, dan tentunya warna daun yang dominan adalah hijau. Dan dengan metoda clustering dapat ditentukan daerah pohon yang mempunyai warna daun hijau, kemudian barulah dengan konsep program komputer ditentukan daerah corn saja atau daerah wheat saja.

Dalam menganalisa data remote sensing, cara pengclustering ini banyak menolong ahli interpretasi antara lain:

- * Memeriksa homogennya suatu data (Checking the homogeneity of data)
- * Menolong dalam mendeteksi garis-garis batas golongan (Aiding in the detection of class boundaries).

Pelaksanaan clustering dapat dikerjakan melalui dua cara utama yaitu :

- * Iterative procedures
- * One-pass procedures

Untuk mengukur jarak titik-titik data dalam suatu cluster dinamakan : Inter - cluster distance atau

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$$

dimana, x dan y adalah dua vector feature
n adalah dimensi ruang feature

$$x = \begin{pmatrix} x_1 \\ x_2 \\ - \\ - \\ - \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ - \\ - \\ - \\ y_n \end{pmatrix}$$

Pengukuran kesebangunan antara dua cluster adalah apa yang dinamakan ; Between - two clusters distance atau

$$D(c_1, c_2) = \sum_{i=1}^n \left[\frac{(r_i(1) - r_i(2))^2}{6_i(1)6_i(2)} \right]^{1/2}$$

dimana, r_i adalah sifar-sifat cluster c_i dengan MEAN r_i
 6_i adalah deviasi standard cluster c_i ($i = 1,2$).

$$r = \begin{pmatrix} r_1 \\ r_2 \\ - \\ - \\ - \\ r_n \end{pmatrix} \quad 6 = \begin{pmatrix} 6_1 \\ 6_2 \\ - \\ - \\ - \\ 6_n \end{pmatrix}$$

10. KESIMPULAN

1. Sistem pengolahan digital dapat digunakan untuk menganalisa data dari multispectral sumber-sumber daya alam dengan cara teknik pengenalan pola dan klasifikasi pola. Dalam teknik pengenalan pola ada dua metoda yang digunakan ;

- * Metoda pengolahan data yang dibimbing
- * Metoda pengolahan data tak dibimbing

2. Dalam metoda pengolahan data dibimbing atau supervised processing ada dua cara yaitu;

- * Metoda statistik (statistical approach)
- * Metoda diskriminan (discriminant approach)

Dalam metoda pengolahan data tak dibimbing atau unsupervised processing digunakan ;

- * Metoda clustering (clustering approach)

3. Pada Supervised processing diperlukan training sample yang ditandai (identified training sample) sedangkan Unsupervised tidak memerlukan suatu training sample yang ditandai, tetapi melalui subgroup atau clustering.

11. DAFTAR PUSTAKA

1. Kamat, D.S, et al : Digital Clustering Techniques Applied to Aerial Photographs for landuse Analysis ISRO - SAC - SN - 01 - 76, 1976 - Indian.

2. A Programming System For Digital Image Processing Of Remotely Sensed Data - ISRO - SAC , 1977 Baugalore, Indian.
3. By: Seymour Shlien, Automatic Interpretation Of ERSTA - A Imagery By Using the Maximum Likelihood DEcision Rule Nev, 1973
4. Kamat D.S., Majumdar K.L. Nasipuri T.K. On Supervised Classification System For Resources Study with Multiband Remotely System For Resources Study with Multiband Remotely Sensed Data - RSA - IP /TN08 /76, 1976 - Indian.

* * * *

SATELIT MAGNETIS

Baru-baru ini Amerika Serikat telah berhasil meluncurkan satelit barunya "MAGSAT" (Magnetic Field Satellite), yang diharapkan mampu melakukan observasi tentang gerakan kerak bumi disamping mengetahui kekayaan tambang.

MAGSAT, dirancang khusus untuk pengukuran medan magnetik dekat permukaan bumi serta pengukuran tidak langsung bentuk-bentuk kerak bumi yang bertalian dengan gempa dan simpanan tambang.

Se lama usianya yang hanya dalam seratus duapuluh hari, Magsat akan melakukan pengukuran kekuatan serta jejak arah dari medan magnetik bumi secara global beserta permukaan medan magnetik.

WESTAR 4

Hughes Aircraft Company merencanakan pelaksanaan pembuatan Satelit Komunikasi Domestik Westar dengan 24 channel transponder. Westar 4 ini akan memiliki dua kali lipat kapasitas yang dimiliki Westar Satellite sebelumnya. Biaya yang diperlukan pada pembuatan satelit Westar 4 ini adalah sebesar 25 juta dollar.

Satelit Westar 4 ini direncanakan diluncurkan pada bulan Januari 1982 dan akan merupakan penerus dari Westar 1 dan 2 yang mendekati masa usai operasinya akhir 1982.

PIONEER 11

Pioneer 11 dari NASA dilaporkan tetap dalam keadaan baik setelah berhasil melintasi Saturnus dengan selamat. Ini adalah rekor terjauh yang pernah dicapai manusia dalam penjelajahan luar angkasa. Pioneer 11 kini tengah melesat keluar dari tata surya, setelah berhasil mengirimkan ke bumi gambar-gambar jarak dekat Saturnus serta memberikan bukti penemuan tentang bulan baru dari Saturnus.

(Space Flight May, June, 1980)